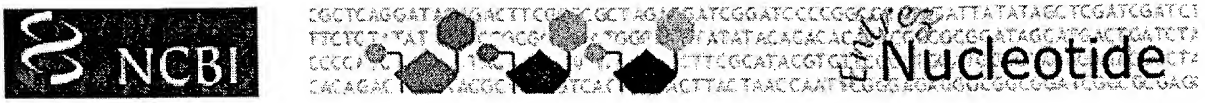
21



1: Z81449. Caenorhabditis el...[gi:1658303]

LOCUS CEC46F11 34841 bp DNA linear INV 02-SEP-1999
DEFINITION Caenorhabditis elegans cosmid C46F11, complete sequence.
ACCESSION Z81449
VERSION Z81449.1 GI:1658303
KEYWORDS HTG; ATP-dependent RNA helicase; Pyridine nucleotide-disulphide oxidoreductases class-I; unc-93.
SOURCE Caenorhabditis elegans
ORGANISM Caenorhabditis elegans
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 34841)
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Shownkeen,R., Smaldon,N., Smith,A., Sonnhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohldman,P.
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans
JOURNAL Nature 368 (6466), 32-38 (1994)
MEDLINE 94150718
REFERENCE 2 (bases 1 to 34841)
AUTHORS Burton,J.
TITLE Direct Submission
JOURNAL Submitted (04-NOV-1996) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu
COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.
For a graphical representation of this sequence and its analysis see:-
http://webace.sanger.ac.uk/cgi-bin/display?db=wormace&class=Sequence &object=C46F11
Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
IMPORTANT: This sequence is not the entire insert of clone C46F11. It may be shorter because we only sequence overlapping sections

```
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            The true left end of clone C46F11 is at 45409 in
            sequence Z48241.
            The true right end of clone C46F11 is at 8174 in
            sequence Z48245.
            The start of this sequence (1..102) overlaps with the end of
            sequence Z48241.
            The end of this sequence (34740..34841) overlaps with the start of
            sequence Z48245.
FEATURES             Location/Qualifiers
     source          1..34841
                     /organism="Caenorhabditis elegans"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:6239"
                     /chromosome="III"
                     /clone="C46F11"
     gene            join(Z48241.1:39319..39501,Z48241.1:42050..42148,
                     Z48241.1:43215..43503,Z48241.1:44316..44431,
                     Z48241.1:44486..44692,Z48241.1:45250..45422,305..539)
                     /gene="C32A3.3"
     CDS             join(Z48241.1:39319..39501,Z48241.1:42050..42148,
                     Z48241.1:43215..43503,Z48241.1:44316..44431,
                     Z48241.1:44486..44692,Z48241.1:45250..45422,305..539)
                     /gene="C32A3.3"
                     /note="similar to coiled coil domains; cDNA EST yk302g12.5
                     comes from this gene; cDNA EST yk365d10.5 comes from this
                     gene; cDNA EST yk461c1.5 comes from this gene"
                     /codon_start=1
                     /protein_id="CAB03766.1"
                     /db_xref="GI:3874997"
                     /db_xref="TrEMBL:Q93317"
                     /translation="MASSELLINRRFDDFILEGCSSDRPSSAVTVLHRPRKKLSAFVS
                     AMAARAASPQSPSKQITVVDVYDLAASIGNDFEKLIDNYGNECVRGIMPKVISALETL
                     EAMAAGNDRENEEIMRLSKAVERLEQEKHQRNQQHLKFEEELEQVEKTYRKDIDDLQQ
                     MVKSLVNENRNLSTTVSSLPNHADSPVSTSMREADLKMLLELKEMSSQQKDEIKALQK
                     DVDTYQCQVENLQNSIEKLIRQNEELLRKNASLQKQGRMIVEEKMEVVRRLEKTEESN
                     IELKKLVKETDRACKDMQLANQDNNEPRFTLSELREVIKEKNILKGRVMELEEELDNF
                     KPGAKKEIMRLDDDDDFDRNPLYFTFYPKNNFFKPKNTTFCSNFWGKKQKKNISSENS
                     KKKNQKTQVSNQICRFLKQTKNREFFVLRYFKKNSIFRKLA"
     gene            2203..6603
                     /gene="unc-93"
     CDS             join(2203..2310,2363..2607,2659..2745,2947..3150,
                     3197..3293,3898..4007,4053..4187,4440..4558,4613..4769,
                     4816..4907,5030..5117,5165..5303,5379..5552,5598..5794,
                     6380..6476,6526..6603)
                     /gene="unc-93"
                     /note="UNC-93 protein"
                     /codon_start=1
                     /protein_id="CAB03760.1"
                     /db_xref="GI:3874991"
                     /db_xref="TrEMBL:Q93380"
                     /translation="MKFQKMGDSNTWDLVGEQQQRKKSRSPSRASRVDSELLVGVENA
                     AELEALGLGKEQLEEEARRHKKQRSKSPALENIRKTSIHLLQKFGAIPKKKDDSVLLF
                     RFHDIPDIPLESLCIRPKEFFEEPKVFSFRDMGREQQKAEVNEKCSYLFRGTSFDHDD
                     HFELPTETGRVPEYDHFCPIHGSRRRLPRNKLVTMQTLMHSVDDEDNEDLAYIYGHDF
                     LAKLVRKKKREMMSGTEKERANKIKRKIMSNLWILSVAFLFLFTAFNGLQNLQTSVNG
                     DLGSDSLVALYLSLAISSLFVPSFMINRLGCKLTFLIAIFVYFLYIVINLRPTYSSMI
                     PASIFCGIAASCIWGAKCAYITEMGIRYASLNFETFQAFNQAKFRFFGYFFMIVHCGQ
                     VVGNMVSSYIFTLSYSQALRGPEDSIYDSCGYQFPKNLSDLTELAESNLARPPQKVYV
```

```
                     AVCLAYLACVIISGMIMSMFLNALAKDARNRKMAQKFNSEIFYLMLKHLINIKFMLLV
                     PLTIFNGLEQAFLVGVYTKAFVGCGLGIWQIGFVMACFGISDAVCSLVFGPLIKLFGR
                     MPLFVFGAVVNLLMIVTLMVWPLNAADTQIFYVVAAMWGMADGVWNTQINGFWVALVG
                     RQSLQFAFTKYRFWESLGIAIGFALIRHVTVEIYLLITFFMLLLGMCGFLAIENFDHI
                     IKFWHHLIHTSCPEKEPLDDRNSDFE"
gene                 complement(6741..8575)
                     /gene="C46F11.3"
CDS                  complement(join(6741..6864,7280..7432,7483..7675,
                     7722..7849,7900..8027,8075..8137,8418..8524,8569..8575))
                     /gene="C46F11.3"
                     /codon_start=1
                     /protein_id="CAB03761.1"
                     /db_xref="GI:3874992"
                     /db_xref="TrEMBL:Q93381"
                     /translation="MTEVEKSRSGRIVKRKVFADFQQLDDDLKSITTEELRESASESK
                     KLRLGPTRRSCRNIPTNYEELKAKTDAGIIPETVYYRASNNSMYIPSITADMLKNDGN
                     DLTVQTRSDKYFKYNRVRQSTKIDYAEPEKVYDIIYAVRKRPILWDQRLICHRNSNLS
                     RRAWDQLDLELGIDEEYPLARRKQIWKSKRDYFVSAVNAANLRKWIYADALEFYRPMI
                     NFRTTFCLRPTVAQPSQSVHDKLVLADKNIQCTPGDKKNVLTFLLKSLYDTGMSSPEM
                     MAEHGLEILKIFKKNAESSNLKLN"
gene                 8804..10704
                     /gene="C46F11.2"
CDS                  join(8804..8838,9157..9411,9467..9824,9875..10346,
                     10403..10704)
                     /gene="C46F11.2"
                     /note="similar to pyridine nucleotide-disulphide
                     oxidoreductases class-I; cDNA EST EMBL:Z14425 comes from
                     this gene; cDNA EST EMBL:C08325 comes from this gene; cDNA
                     EST EMBL:C10026 comes from this gene; cDNA EST yk505a7.3
                     comes from this gene; cDNA EST yk489g8.3 comes from this
                     gene; cDNA EST yk489g8.5 comes from this gene; cDNA EST
                     yk483c8.3 comes from this gene; cDNA EST yk483c8.5 comes
                     from this gene; cDNA EST yk437a9.3 comes from this gene;
                     cDNA EST yk437a9.5 comes from this gene; cDNA EST
                     yk387c6.3 comes from this gene; cDNA EST yk387c6.5 comes
                     from this gene; cDNA EST yk456b3.3 comes from this gene;
                     cDNA EST yk456b3.5 comes from this gene; cDNA EST
                     yk428g11.3 comes from this gene; cDNA EST yk428g11.5 comes
                     from this gene; cDNA EST yk426e9.3 comes from this gene;
                     cDNA EST yk367a12.3 comes from this gene; cDNA EST
                     yk298a12.3 comes from this gene; cDNA EST yk298a12.5 comes
                     from this gene; cDNA EST yk305a1.3 comes from this gene;
                     cDNA EST yk292b1.3 comes from this gene; cDNA EST
                     yk292b1.5 comes from this gene; cDNA EST yk289c12.3 comes
                     from this gene; cDNA EST yk289c12.5 comes from this gene;
                     cDNA EST yk247a4.3 comes from this gene; cDNA EST
                     yk247a4.5 comes from this gene; cDNA EST yk228b6.3 comes
                     from this gene; cDNA EST yk228b6.5 comes from this gene;
                     cDNA EST yk196f4.3 comes from this gene; cDNA EST
                     yk196f4.5 comes from this gene; cDNA EST yk428g7.5 comes
                     from this gene; cDNA EST yk510b6.3 comes from this gene;
                     cDNA EST yk521b3.3 comes from this gene; cDNA EST
                     yk614f9.3 comes from this gene; cDNA EST yk640g10.3 comes
                     from this gene; cDNA EST yk642c4.3 comes from this gene"
                     /codon_start=1
                     /protein_id="CAB03763.1"
                     /db_xref="GI:3874994"
                     /db_xref="TrEMBL:Q93379"
                     /translation="MLRFRCILSTSRSIMSGVKEFDYLVIGGGSGGIASARRAREFGV
                     SVGLIESGRLGGTCVNVGCVPKKVMYNCSLHAEFIRDHADYGFDVTLNKFDWKVIKKS
```

```
                RDEYIKRLNGLYESGLKGSSVEYIRGRATFAEDGTVEVNGAKYRGKNTLIAVGGKPTI
                PNIKGAEHGIDSDGFFDLEDLPSRTVVVGAGYIAVEIAGVLANLGSDTHLLIRYDKVL
                RTFDKMLSDELTADMDEETNPLHLHKNTQVTEVIKGDDGLLTIKTTTGVIEKVQTLIW
                AIGRDPLTKELNLERVGVKTDKSGHIIVDEYQNTSAPGILSVGDDTGKFLLTPVAIAA
                GRRLSHRLFNGETDNKLTYENIATVVFSHPLIGTVGLTEAEAVEKYGKDEVTLYKSRF
                NPMLFAVTKHKEKAAMKLVCVGKDEKVVGVHVFGVGSDEMLQGFAVAVTMGATKKQFD
                QTVAIHPTSAEELVTMRGGVKPE"
gene            complement(10910..15329)
                /gene="C46F11.5a"
gene            complement(10910..15329)
                /gene="C46F11.5b"
CDS             complement(join(10910..10987,11038..11129,11176..11304,
                11836..11956,11998..12157,12229..12386,12779..12975,
                13086..13136,13628..13804,14097..14184,14241..14394,
                15217..15329))
                /gene="C46F11.5a"
                /note="cDNA EST CEMSA24F comes from this gene; cDNA EST
                EMBL:M79505 comes from this gene; cDNA EST EMBL:D66149
                comes from this gene; cDNA EST EMBL:D69828 comes from this
                gene; cDNA EST EMBL:C11785 comes from this gene; cDNA EST
                EMBL:C10226 comes from this gene; cDNA EST yk401g7.3 comes
                from this gene; cDNA EST yk401g7.5 comes from this gene;
                cDNA EST yk327a4.5 comes from this gene; cDNA EST
                yk469b4.5 comes from this gene"
                /codon_start=1
                /protein_id="CAB03762.1"
                /db_xref="GI:3874993"
                /db_xref="TrEMBL:Q93383"
                /translation="MERIRESGRRLHQVYRETIEWRNPLTGLYLMAVNSAFWIGVIYC
                DPRIQEALLATASAGVFAWDILLSSSNDRSIITHLLMWPFQSIFRTLSVGGSVYSIHL
                LRAEQVRLACYAAYATLACLLVNPVWEHNEVNSKIASATRRTVSWFGSWIQYLIITPV
                VTVYEYTKYIVLFRWVPQKINSKQKCFKFFVKNSPLIAYVKHFIFNFRESCRELVNGI
                KNWIHVAIVERTKRVAQRIRQFLRYWFCAEWWPTLKEWLKLNIGVPLRYLFDQLCFVF
                VYIFCAHWFPPLWKFSVKQLKSLGALAYKHIWVPVKGFLLCQFEKLRCWLRDTLHRIA
                ISVRDSILWPICCLMVEVGKQVSIFVYHLLLEPIVNYLHSRYKIIETSALIHVIGPVC
                ETVIDHIPEKNPFCEESDVELEGFLPEVNDETDLDENHVDDDDNEDIQSELSSSPIPE
                EEFEFERGLQFSAINGSESSDEEFDLDAPKKTIRRRRREPKKSEVAADDEYELLE"
CDS             complement(join(10910..10987,11035..11129,11176..11304,
                11836..11956,11998..12157,12229..12386,12779..12975,
                13086..13136,13628..13804,14097..14184,14241..14394,
                15217..15329))
                /gene="C46F11.5b"
                /codon_start=1
                /protein_id="CAB03764.1"
                /db_xref="GI:3874995"
                /db_xref="TrEMBL:O62100"
                /translation="MERIRESGRRLHQVYRETIEWRNPLTGLYLMAVNSAFWIGVIYC
                DPRIQEALLATASAGVFAWDILLSSSNDRSIITHLLMWPFQSIFRTLSVGGSVYSIHL
                LRAEQVRLACYAAYATLACLLVNPVWEHNEVNSKIASATRRTVSWFGSWIQYLIITPV
                VTVYEYTKYIVLFRWVPQKINSKQKCFKFFVKNSPLIAYVKHFIFNFRESCRELVNGI
                KNWIHVAIVERTKRVAQRIRQFLRYWFCAEWWPTLKEWLKLNIGVPLRYLFDQLCFVF
                VYIFCAHWFPPLWKFSVKQLKSLGALAYKHIWVPVKGFLLCQFEKLRCWLRDTLHRIA
                ISVRDSILWPICCLMVEVGKQVSIFVYHLLLEPIVNYLHSRYKIIETSALIHVIGPVC
                ETVIDHIPEKNPFCEESDVELEGFLPEVNDETDLDENHVDDDDNEDIQSELSSSPIPE
                EEFEFERGLQFSAINGSESSDEEFDLDAPKVKTIRRRRREPKKSEVAADDEYELLE"
gene            complement(15640..22126)
                /gene="C46F11.4"
CDS             complement(join(15640..15928,16934..17369,18285..18464,
                18515..18754,19660..19759,19808..19939,19992..20168,
                20408..20579,21186..21365,21437..21688,21741..21811,
```

```
                     21875..22002,22048..22126))
                     /gene="C46F11.4"
                     /note="similar to ATP-dependent RNA helicase; cDNA EST
                     EMBL:D37484 comes from this gene; cDNA EST EMBL:D64936
                     comes from this gene; cDNA EST yk371b5.3 comes from this
                     gene; cDNA EST yk371b5.5 comes from this gene; cDNA EST
                     yk201a4.5 comes from this gene; cDNA EST yk378c11.5 comes
                     from this gene; cDNA EST yk484b11.5 comes from this gene"
                     /codon_start=1
                     /protein_id="CAB03765.1"
                     /db_xref="GI:3874996"
                     /db_xref="TrEMBL:Q93382"
                     /translation="MWRGKYSGGYQNRFVSAKSSGTSSLSDGKSMNNVPPPAALTVGN
                     APVPPPNRNILLSNRKEQEMKKEQEYANLMAYGSTSEKQSRRKRIYDDEYLEGGSDDE
                     KPASSSSKLAADDDDEEDELDAFMAGIEKQASSDKKVSEQKEKDRKEGKDTEDPSKKG
                     LGRADIDEEDMQESLFKFMEEYKEKHENDDEQLEYDEDGNIIWSWKKVIDPLPDIDHS
                     QIQYQKFNKNFYEEHEDIKRLHYMDVIRLQNTMNLRVGGLKPPRPVCSFAHFSFDKLL
                     MEAIRKSEYEQPTPIQAMAIPSALSGRDVLGIAKTGSGKTAAYLWPAIVHIMDQPDLK
                     AGEGPVAVIVVPTRELAIQVFQEAKKFCKVYNINPICAYGGGSKWEQSNELQNEGAEM
                     VVCTPGRIIDLVKMGATNFLRTTFLVFDEADRMFDMGFEAQVKSISDHVRPDRQCLMF
                     SATFKQKVERLARDALVDPVRIVQGEVGEANADIEQKVFVMQNQDVKLHWLIRNLVEF
                     ASLGKVLIFVTKKLDSEDVAKKLKMKDFDIVLLHGDMLQAERNENLLKFRKKSQILVA
                     TDVAARGLDISEIRTVINFDMARDIDTHVHRIGRTGRAGHKGTAYTLVTEKDIEMVGH
                     LVKNLESVSQEVPKPLMDLAMKSSWFRGQRAGNGGPSTGTQTRGRMGLGYTPKVRQVG
                     GGATGAQFDPLKEQKSCRGGGNQTVDGMIRNAQSFATSSSSSSSSSSGPASGANRAQM
                     LKSAFQKSFQTTFQRPTNESSNLPQQVASDPRPEWKKKVDELNAKIAQQQASSASGGS
                     SSSSSSSKRSRWE"
ORIGIN
        1 gatcggaatc cattgtacgt tgtttttga aatttttta taacaaatta ttgaaaaaaa
       61 aactagaaat tcagaaaatt gaaattggaa ttaacatgat cattttaag acttttaaa
      121 gaagattgaa aattataaaa atttattttg aaaaaaaaaa tttttaattt gctctaaaat
      181 actcaaaaaa aaacggaatt tttaatacta ctactttaaa aatttgtaaa ttaaaaaata
      241 acataattca actgaaaaaa cggctaaagc tagcttcttt acctaaattt tcagatattt
      301 taagatattt tacattttat ccgaaaaata atttttttaa accaaaaaac acaacattct
      361 gttctaattt ttggggcaaa aaacaaaaaa aaaacatttc ttccgaaaat tcgaaaaaaa
      421 aaaaccaaaa aacacaagtt tccaaccaaa tttgtcgatt tttgaaacaa acgaaaaatc
      481 gagaattttt cgttttacga tatttcaaaa aaaattcgat ttttcgaaaa ttagcttaaa
      541 ttttggccat tgttcttctt tttagtataa tttttttttg ttttttctca tcaatataca
      601 caaaattaga tttaagaaaa ataatatttt atagttaatt ttctgaaaaa ttttttccgat
      661 tgttcgattt tccctcaaaa aatcaaaaaa caccaaaaaa aaaaccattt caaaatatgg
      721 aaaatgttat taatttctgg ctggagacaa ctttttttctc aaaatttttcc taaaagtaac
      781 cccaaaactt ttttttcaga caaccagaca cctcttcggc ccaaagtctc ggcgttgacg
      841 gtgacgattt accggtttat ggtccaatgc caaaagaacc cgacgagaag cttcatccat
      901 ggaaatatga gagaaaagat tccggagttc gaaaattgtg agtttatcct tcttcccatt
      961 tcaaattttt aaaccatgta aaaagtattc agtttccgct tcttcaaaga ctttggagcg
     1021 acggcgtccc cccgccgagg atcttctgca aacgcgtctc ctcgactacc ggctcgtgcg
     1081 atcccaccat ccctcaattg atgatcatgg ctccgatgct tttgccagtt catattttct
     1141 tttctgcaat ggttttcatg tttcagtgtc tgtgacaccc ctccattctc tctaatcatc
     1201 acatttttg atctcactac tagtcaaacc ccatttcagc ttcaaaagta ttttttttt
     1261 tttaaatttg tacgacatgt ctctcactct ctgcctgctc cttgatgatt atgatcgact
     1321 atcgatcgat tttaggctca atttcgcttc caaaaaaaat ctattcattt tccccgtgtt
     1381 ttacccttcc ctgtctctta ttttttgctc ttcatgtggc atatatatgt aattatgtgt
     1441 gttgtcatac tttttgtcat ctttttttgaa atatcgaaat taacaaagaa tttctcacaa
     1501 aaattcacag ttgttttcaa ttttccacac acaaaacatt tttcaaaact atcacggaaa
     1561 gcttctgctg ctgcgtggtc agagattttc ggagcaaaag agcaatgtgt aacactggcg
     1621 ctctggagtg cgggaaggaa gaaaaacgaa aattcagaca gctgttgagc tgaacaattc
     1681 tttgcgttga aaatactgat ctggttcgca ttttttacgg atttttatta tgaattttgg
     1741 agatttttgg acatttctgc actattttac cggaaaactt gctatttcaa atcttaaaag
     1801 agcacagaaa tttaaattta tattaatatg cattttcact ggagtttgaa cgcttttta
     1861 gttatttag tcattaaaaa actatggatt tgagctgaca ttttttgaaa ttcgaaatta
```

```
1921 atttgtggct tttctgtttt ttttttgaaa aatgttttac attaaaataa ttaaataaac
1981 taaatttcac atttctgcta atattttatg aaaactttat gacaaaaacc ctgtaaaaat
2041 gcgaaattta gacttttacg atataaaaat ttaaaaaaac tggtttgtcc gagaggagta
2101 cacgccggat gcgctttgcg acttttgcgc ggtaattcaa atttgaattt tccgcatagt
2161 tgaatttagc gtattcttag ggcaaattaa cagaaaaaat gtatgaaatt tcagaaaatg
2221 ggtgattcaa acacctggga ccttgttgga gagcagcagc aacggaagaa aagtagatca
2281 ccgtcaagag caagccgtgt ggattcggaa gtaattttaa gcgcattcaa tatgtactaa
2341 ttaaaggtgc atacactttc agctacttgt cggtgttgaa aacgcggctg aacttgaagc
2401 tcttggccta ggaaaagagc agctggaaga agaagccaga cggcacaaaa agcagcggag
2461 caagtcgccg gctctcgaga atattcgaaa aacttcgatt catcttttgc aaaaatttgg
2521 agcgattcca aaaaagaagg atgatagtgt tttacttttt agatttcatg atattcctga
2581 tattcccttg gaatctttat gcatcaggtg agaaagaagt tgcaaaattt attttgaaaa
2641 ttcgaaaaat aatttcagac cgaaagagtt tttcgaggag cccaaagtat tttcctttcg
2701 agacatgggc agggaacagc agaaggcaga ggttaatgag aagtggtaag gggtttttga
2761 atttgcgcgt caaattcggt gtttgcattt tccaagctcg agcgacaatg gccgaggatt
2821 tttctcggcc atatagaaaa accgatctct tgcgtttcta agttgctaca gccgcgaaaa
2881 aaaacatcgg tggccgagtt ttccaatttt cccggccacg aatcaaaatt cggaatttat
2941 tttcagctcc tacttattcc gtggcacctc attcgaccac gacgatcatt tcgaacttcc
3001 cactgaaaca ggccgtgttc cagaatatga tcatttttgc ccgattcatg gatccaggcg
3061 aagacttcca cgtaataaat tggtcacaat gcaaactctc atgcattcgg tagatgatga
3121 agataatgaa gatttggctt atatttatgg gttagttttc agagaaattt gaatgaaaat
3181 tttaatcggt tttcaggcac gactttctgg cgaaattggt taggaagaag aagcgagaaa
3241 tgatgagtgg aacggaaaaa gaaagagcta ataagattaa aaggaaaatt atggtaattt
3301 tcaaaacgtg gctgcgaaat gggaaaacta ggccaccaca gaaaactagg ccaccagaga
3361 aaactaggcc gctagaaaga taaactaggc cactagaacg ataaactaga ccgcttgaaa
3421 aactaggcca cgatcaaaaa agctactact attgaaaaaa aaaactagac cactggacaa
3481 aaagctaggc cgttcgaaaa aagacactag aaaataacta ggcagttttt aattaatttc
3541 attaattaat tgcaattaat caagtttcaa aaaatgcgaa aactcggcca ccgacttttt
3601 tcgcggccgc gtggttccaa actatgaaaa aaataacagc aaagtcggtg gcctagtttt
3661 attgtctctt ccaaatgaca catacgattg taaaaatggc tgggaaaatc agtggcctag
3721 tttttttcta gtggcctata atttctctag cggcctagtc ttgattgcta gtggcctagt
3781 ttttttttt ttctggtggc cgagttttcc tttcaaatta ggatacttag gatatatagt
3841 ggccgagttt cctgttttcg ccacttcaac tagaccaact aatcgaatca ttttcagtca
3901 aatctctgga ttctgtccgt ggcttttctc ttcctgttca ctgcatttaa cggactccaa
3961 aatcttcaaa cttccgtcaa tggagatctc ggatctgata gtttagtgta agaacgtttt
4021 gaatttcaat ttgaaaaata ataatttttc agagccctat atttatcact tgccatatca
4081 tcattatttg ttccatcatt tatgataaat cgcctcgggt gcaaattaac attccttatt
4141 gcaatcttcg tgtattttct atacattgtc attaatttga ggccgacgta agtttttgcg
4201 ttttttgagc gcttgtaact cagaaactaa aatagttagc aaaaaactgt caactaacaa
4261 aatgttaatc gatttttgat ctacattttg ctagttgaca acttttttgat atctttcacc
4321 tataagaagt tatgtgcatt tgaagagcta gagatataaa aaatttattg ccatttattt
4381 ccgtttagaa ctttttgctg tttagcattg ttagttctaa gctacgtggc tttttccaga
4441 tactcttcta tgatccctgc ttcaatattc tgtggaatcg cggcttcatg tatctgggga
4501 gccaaatgcg cgtacattac agaaatggga attcgatatg ctagcctgaa ctttgaaagt
4561 cagactactg ttattgttag gtaatgggct gattttctaa acctttttca agcttttcaa
4621 gcttttaatc aagccaaatt cagatttttc ggatattttt tcatgattgt acactgcgga
4681 caagtcgtcg gaaatatggt atcctcttat atttttacac tgtcctattc gcaagcccta
4741 cgcggtcctg aagattctat atacgatagg tttgccgaaa tttttttttg aagttttaaa
4801 atttcaattt tccagctgtg gctaccagtt tccaaagaat ttatcagatc ttaccgagtt
4861 ggcggagagc aatcttgctc ggccaccaca gaaagtttat gtagcaggtg agcctacata
4921 cctacatacc tacataccta catacctaca tatctacaaa cctacatacc tacataccta
4981 catacctaca tacctaccta cctataccca attttttttt aaatttcagt ttgtctggca
5041 tacctcgctt gtgtaataat cagtggaatg ataatgagca tgttcctgaa tgctttggcc
5101 aaagatgctc ggaacaggtt tgcctacttc cagtcttctc ttaaatttgg ctctgaggtt
5161 tcagaaaaat ggctcaaaag ttcaactcgg aaattttcta cctcatgctc aaacatttaa
5221 tcaatataaa gttcatgttg ttggtgccgt tgacgatttt caatggtttg gagcaagctt
5281 ttctagttgg agtttataca aaggttagtg ttacgtggcc gttaagaggg taaactcggc
5341 cacgtcacta aatcgcaaaa ctgaaaattt tcttccaggc attcgtcgga tgtggccttg
5401 gaatttggca aataggcttt gtgatggcct gcttcggaat ctccgacgca gtctgttccc
5461 tagtcttcgg gccactgatc aagcttttcg gtcggatgcc tctttttgtt ttcggagcag
```

```
5521 tagtgaatct attaatgatt gttacactta tggtatgatg caccatattt gacgcgcaat
5581 ttcatatctc ctctcaggtt tggccactaa atgcagccga cactcaaata ttctacgtag
5641 tcgctgcaat gtggggtatg gcagatggtg tctggaatac tcagattaat ggattttggg
5701 ttgcacttgt tggaagacag tctctacagt ttgcattcac taaatatcga ttttgggaat
5761 ctcttggaat tgcgattgga tttgcgttga ttcggtgaat ttttttttg gcttttgccg
5821 cattaaaaat ttatggggac gcatccattt tgtaatcgat gtggaaacgc gctccacggg
5881 caattgaaaa cgctccgccc cctacagtag ggtctcgtta ggtagttgtg gtgggacctt
5941 gtaaattcaa acttttcaa ttagtttcgc cgatttccat gcatttttcg ttttttttgt
6001 tatattttcc gttctttgta aggatttttt cgccgaaatt gatgaaataa agtggaattt
6061 aaaaattgaa ctatttttta aactaaaaac gtatttttaa acataattag tggaaaaaat
6121 gacgaaaaac atttttagaa acacaaaaaa catttaaaac gttaaaattc gcgttttggt
6181 ccccaatacc taacgagacc cagcttttgg gggcaggaca tttgcattta gccgtggagc
6241 acgcttgcac ctcaatctga ctattcgcat tttttgcgtc tccataataa tgttggccgt
6301 gaaaaaaaca ggtggccgag ttttctttta aatttcccgg ccacgctaga tttaacttga
6361 aaaataattg aattttcaga cacgtcacag ttgaaatata tttgctgata acattcttca
6421 tgctactttt aggaatgtgc ggatttttag caattgaaaa tttcgatcac attattgtaa
6481 ggctgattcg aagaaattaa ttttttctaa ttttaaattt tccagaaatt ttggcatcac
6541 ttaattcaca cgtcatgtcc agagaaggaa ccgttggatg atagaaattc agattttgaa
6601 tgaaaataat taaaaatcaa aaaaatcagt taataaattc gtactttatg ttttattaaa
6661 aattcgaaaa gtcgagaagt ataacaaaat gcagtaaaat gagaaatatg tgataaaaat
6721 taataaaacg tggcgggtgg ctaattgagt ttcaaattgc tagactctgc gtttttcttg
6781 aaaattttca gaatttccag tccatgctca gccatcattt cgggactcga cattccggtg
6841 tcgtagagag attttagaag aaatctggaa aaatttaata ggattaagcc gtatgtgtcg
6901 atttacgcaa gtcgtgtact cctcgagaag aagaagaaca cagttttcgg ctaaaattct
6961 tgatttggaa ctgtttgagc taatttttc tgtattcgag tttagagcac cgttttgtag
7021 ttacagtagt tttcgtgacg ggacccaaat tggtttagca cccgtcgagc tgtattcaaa
7081 attcctgctg aaaatagaaa aaggattttt taatcatgtt ttctatactt tttttgtttt
7141 ttgaagtttt ttacagcaaa agaattggcg aagatggttc aaaatagatc aaaatatacc
7201 caaaagttgt gaatttttaa aaataaaatg caaacttctc ttcgaggagt acacttttta
7261 tgttgggaat tttcctcacg tcagaacatt tttcttatca ccaggtgtgc attggatatt
7321 cttatcagcc aatactaatt tatcatgaac actttgagat ggctgagcga cggttggacg
7381 gagacagaat gtcgttcgga aatttatcat tggacggtag aactctaatg cactgaaaaa
7441 attattattt tcatttcatt attttcaatt ttccagccgt actctgcata aatccatttt
7501 ctcagattgg cggcgttcac cgctgatacg aaataatctc ttttactctt ccagatttgt
7561 ttacgccgag ccaatggata ctcttcatct attccaagtt ccaagtcgag ttgatcccaa
7621 gcacggcgag acaaattcga attccgatga caaatcaggc gttgatccca taggactaaa
7681 aagggaaatt tttggaaaaa cttcttgatt tcctgactta ctaggccttt tccgaactgc
7741 atatataata tcgtacactt tttcaggctc cgcgtaatcg atttttgtag attggcggac
7801 ccgattgtac ttaaaatact tgtctgagcg tgtttgcaca gtcaaatcac tgaaaaaagg
7861 ataaaaaaga aatctgataa taaatgtttt aattcatact ttccatcatt cttgagcata
7921 tcggctgtaa ttgaaggaat gtacatcgaa ttgttggatg ctcgataata aactgtttcg
7981 ggaataattc cggcatcagt tttcgctttt aattcttcgt aatttgtctg aaatctttga
8041 atagaattga aacgaatatt ataaaaagct gtaccggaat attgcgacaa gatcttctag
8101 ttggtccaag acgaagtttc ttcgattcgc tggcgctctg aaaaaatgtt tgatttgtca
8161 tcatttttc tagttttcag tgaaacttt gtcatcattt attagttttt tcttgtgcct
8221 gattgaattg attgaaacta catataaata atcactattt taattttatt gcccactcgc
8281 gctctttttg gttttttct aataaaattg tctttcttct ctgcgcactc cactattgac
8341 caaatcagtg tttctttcta tttgtcattg ccccgacact gcgaagaacg ttcgcgattg
8401 ttctttgtct tttttacctc tcttaattcc tccgttgtaa ttgatttcaa gtcgtcatcg
8461 agttgttgaa aatcggcaaa cacttttcgc ttcactattc ggccgcttcg agatttttca
8521 acttctgaaa aattgagaga ttcatataga ccatttttca acatttacct gtcatttttt
8581 attccagcgt gcgaatccgt ttggaaaatt gagcggaatt tttaatttcc aaaaaaactt
8641 ctaaatttca aattttaaac tcttcgcgcg ctgaaaaagg ggcggagttt tcaattcgca
8701 aaatgggcgc gagctatttg caacttgggc gcacaattta gtccttgatg tttgtttatt
8761 ttttgctgat caaatgtcta aaaacttctg tgaacccttc tgaatgctcc gatttcgctg
8821 cattttgagc acttcgaggt attaaattgt tacttttttgg tttaaaaaaa gtttttgtag
8881 tctttgtgga gtcaaatttc cataaaacct ctcgaatttc ggaagaaggg cggtcacatg
8941 tttgtagtcg cgctttgcta tctttgtaat ttgtcatcgt ttttgcgtca gtcgattttg
9001 agtatttcat tcgaaaaata aactttaaag tccacttcgc gactgaaaag aaataaagct
9061 ttgatttttc tgcgaatttt caatcgcgcc tctctttcga tgttcttaca agttctgtta
```

```
 9121 ttttccaact tttaattatt aaaattcgag ttccaggtcg atcatgtctg gcgtcaagga
 9181 gttcgactat ttggtgattg gtggaggatc tggaggtatc gcttctgctc ggcgtgctcg
 9241 tgaattcgga gtttccgtcg ggctcatcga atctggacgt ctcggaggaa cttgcgtcaa
 9301 tgttggatgt gttccgaaga aagttatgta caattgctct ctccacgccg aattcatccg
 9361 ggaccacgct gattacggat ttgatgtgac gcttaacaag tttgattgga agtgagtaat
 9421 ttttaattag tttaatcaaa taattcttaa ttgttttgtt ttttagagtg atcaaaaaat
 9481 cgcgagatga gtacatcaaa cgactcaatg gtctttatga gagtggactg aaaggatcct
 9541 cagtcgaata tattcgagga cgtgcaactt ttgcagaaga cggaactgtt gaggtcaacg
 9601 gagcgaagta tcgtggaaag aacacgctca tcgctgttgg aggaaagcca accattccaa
 9661 acatcaaagg agccgaacac ggtattgatt ccgacggatt cttcgatctc gaagatctgc
 9721 caagccggac cgttgtcgtt ggagctggtt atattgccgt agagattgct ggagttctcg
 9781 caaatcttgg ttcagacacg catcttctta ttcgttacga taaggtacca tgaggaacat
 9841 ggaaactttt caagtttaac taaataattt tcaggttctc cgcacatttg acaaaatgct
 9901 cagcgatgaa cttactgctg atatggacga ggaaacgaat ccacttcact tgcacaaaaa
 9961 tactcaagtc acagaagtaa tcaaaggaga tgatggtctt ttgacaatca agacaacgac
10021 tggagtcatc gaaaaggttc agactttgat ttgggccatc ggaagagatc cactgacaaa
10081 agagcttaat ctggaacgtg ttggagtgaa aaccgacaaa tctgggcata ttattgtcga
10141 tgagtaccag aacacatctg ctccaggaat tttgtctgtt ggagatgata ctggaaagtt
10201 ccttctcaca ccagtcgcaa ttgctgccgg acgtcgactc tctcatcgtc tgttcaatgg
10261 tgaaactgat aataagttga cttatgagaa cattgccaca gtggttttca gtcatccact
10321 tattggaacc gtcggactca cggaaggtat aatatcttta aaatttaatt ctttgtgtca
10381 aatttaatcc ccaaaatttc agctgaggcc gttgaaaagt acggaaaaga cgaagttacc
10441 ctctacaaat ctcgcttcaa tccaatgttg ttcgcagtca ccaagcacaa ggagaaggcc
10501 gcgatgaagc ttgtttgtgt cggaaaagac gagaaagtcg tcggagtcca tgttttcgga
10561 gttggatccg atgagatgct ccaaggattc gctgttgctg tcacaatggg cgccacgaag
10621 aaacaattcg atcagactgt cgcgattcac ccaacttctg ccgaagagct cgttactatg
10681 cgaggaggtg tgaagccgga ataatttttt atttatttat tacattttta aatttggatc
10741 aatttctttt gactttcttg tagataaaac attgcaaaat ttaaaatgca tggcaaaaaa
10801 attatttaca ggttatgcaa aaaaatctaa ttatggtgag tgctacatat taacccgatg
10861 ggtagatttt ggtaggaaag gattttcaga attaaaactg caattgagac tattcgagaa
10921 gctcatattc gtcatcggcg gccacttctg atttttagg ttctcgtcgc ctccgtcgga
10981 tggttttctg caaaatttgg agttttttgg aaattattaa aaagaagaaa aaactacttt
11041 cggagcatcc aaatcgaatt cctcatcact ggattccgat ccattgattg ctgagaattg
11101 aagacctcgt tcgaattcga attcttcttc tgaaatttat taactaaaat aatttaaaga
11161 gtctctgaaa ataacctgga atcggcgaag aagacaattc cgattgaata tcttcattat
11221 catcatcatc cacatgattc tcatcgagat cggtttcatc gttcacttct ggaagaaatc
11281 cttcgagttc cacatcgctt tcttctgaaa aatcgatatt taaagtaaaa aattgcttaa
11341 caaagtgaaa acgcgctcca atgccaactt tatatagtcc accctattct cgtgggtctc
11401 gcctggaatt tctagctttt tgtcgttttt taaaacattt ttcggtagtt ttaatcattt
11461 ctatgtcaaa attggttaat taaaaatttt ttagattcat ttttacatgg aaatcagaaa
11521 aatagaggaa aagtgtaaaa agcaattttt accattttaa aaaattttgt ctactccttc
11581 gtttctcgta aaaaagtttt tttttccgaa ttaattaact cgttttcaat aatttttaga
11641 ttaattcaaa caggaaaaat agggttgaaa aaaaaatttt tgtctaaaaa gccctttta
11701 tttttatttt aatgtaaaaa atgatcaaaa aatcgctttt ccgtagtgta aatttccgtt
11761 tttttccaga cgagacccaa ataaaagggg cagatcgatg ggagtttgca ttggagcgcg
11821 tttcataaga cgcaccacaa aatggatttt tctcgggaat atgatcgata actgtctcac
11881 agactgggcc aattacgtgg attaatgcac ttgtctcgat gattttatat cgactatgga
11941 gataatttac aatgggctga aaaattgaat ttttaaaata aaattaaaaa aaattacctc
12001 caggagcaag tggtacacaa aaatcgagac ctgcttcccc acctcaacca tcagacaaca
12061 aattggccag agaatcgagt ctcgcacgga aattgcaatt cgatgaaggg tgtcgcggag
12121 ccagcaacgg agtttttcaa attggcacaa taggaatctg gaatttttt gttttttctc
12181 aattacttta aaagtataga ataatgaacc tagaactctc aaactaaccc cttaaccgga
12241 acccaaatat gcttataagc caatgctcca agtgatttca actgcttaac actgaatttc
12301 cagagtggtg ggaaccagtg agcacagaaa atatagacaa acacgaaaca aagttgatcg
12361 aaaaggtagc ggaggggaac accgatctga aaaatcttgt gctaatctgt ataattctat
12421 tgtagtaatg ggggaacaca ggccaaacaa tactcaaatt ggcccccaaa agaacggaat
12481 ttattagcct ttttcaaaat tttgtcgagt ttttttgctt attattctcg tttatctgac
12541 gatttttcgt tgatttaagt acattttgag ctgattattc aaagaatcaa ataaaattag
12601 ggtttgaggc cattataaac atagaaaaga aggcaataga aaattttcta acaatttgag
12661 gaaattatga acaaaaaaaa taggaacaat gaaaaattat gacggcattg aatgagcaaa
```

```
12721 aaagaagcag ggaaaatttc ataatagaac tgataaaaaa cgaaattaaa atgcttacat
12781 tcaatttcag ccactctttc aacgtcggcc accattctgc acaaaaccaa tatctgagga
12841 attgacgaat tctctgagca actcgttttg ttcgttctac gatcgccacg tggatccagt
12901 ttttgatacc attgacgagt tcacgacatg attcacggaa attgaagata aaatgcttta
12961 cgtaggcgat tagagctgta aaaagagagg gggttttcga ttttttaaag aaaatatttc
13021 aaatttttg aaaaattttt tttcaataaa ttatgttttt ttttcctgaa aattaatagc
13081 tttacgggaa tttttcacaa aaaatttgaa acatttttgc ttcgaattga ttttttctgg
13141 aaaactggca ataaatagat tcgattttcg atttttcgga aaacttttaa tttttaaaaa
13201 tgttcataaa taaataattt taaaaatccc aaaaaaatat agtgtttcgg aattttaaat
13261 taaaaaattg gaaaatctaa aagttaccgt atttttttga tttttcggga agagaaaatt
13321 ccaatttcca ttaatattgt caattttttt cataaaaaaa tattcttaag tttctcatat
13381 ttccggaatt ggtttttgaa aacaaaaaat ttaatgtttt tatttggaaa atttttttat
13441 tttttttatt ttcaattttt caaataatgt acttattgcc aattttaaaa tttatttcaa
13501 tgaaattttc atttttcttt agataacaaa catttacaaa atttcgtaca aaaatcatag
13561 ttttttccat tttccgattt tctccgaaaa ttcaatattt ttgtgaaaaa tatttttaaa
13621 aacttacgtg gaacccatcg gaagagcaca atatatttag tatactcata tacagtaaca
13681 actggagtaa taataagata ttgaatccaa cttccaaacc aggaaactgt ccgccgcgtt
13741 gctgacgcta ttttcgagtt gacttcgttg tgctcccaca ctggatttac aaggagacac
13801 gctactgaaa atggaaaatt tttttttgtt ttttttaatc taagagactt agtattgcgc
13861 gacatatctc gtaacgaaat ctacagtaag tgactttaag tgactactgt agcgcctgtg
13921 tcgatttacg atttacgggc tcgagtttcg aaatgaattg gaatcttttg aaaatcgaca
13981 caagcactac agtagtaatt caaagagtta ctgtagtttt cgctacgaga tattttgcgc
14041 gtcaaatatg ttgcacaata cgcattctca aaatttcatg tttttataat acttacaagt
14101 tgcataagca gcgtaacaag ccaaccgaac ttgctcagca cgtaataaat gaatgctgta
14161 gactgatcca ccgactgaca gcgtctataa ataaacatta atttttagtc ttttataatc
14221 ggattaaaaa atggatttac tctgaatata gattgaaatg gccacattaa aagatgtgtg
14281 atgattgaac gatcatttga agaggaaagt agaatatccc aggcaaaaac tccggcggat
14341 gcagttgcta gaagagcttc ttggatacga ggatcacagt agatcacacc gatcctgaaa
14401 gaaagaggat ttttttaaaa tgaaattggt tgaattttag gatttttaag tgcaaaattt
14461 tggttttttt ttcagtttgt tgtggttttt tcagtgaaaa tatagatttt ttgcttagtt
14521 atcgcagaaa tgatgatttt tctttgaaaa tttggtgttt ttttattcat tgcaaaaatt
14581 ggtttttttt tttcattttt ttaaaatttt ctatgaaatt tttttaaaat tatttattgc
14641 agaaatttga tgttttccgc cagattttcg cagagttcca tatttttga ttaattaaca
14701 aattcacgtg tgtattaccc aaacgcaaaa gtttacggta aaaaacggtc tcggcacgac
14761 aaatttttgt taaatgcgaa gagttgtgcg cctttaaaga gtactgtagt ttccaaagaa
14821 aatttcatcc atttttcata gttttttgaa cgatttctta aaaacttgtg tttttatagt
14881 tttatttaag ttgtaaaaaa acaatttttc tttttttttt aattgaaaaa ccataaaaat
14941 cgatgaaaat ttctcagaat tgtttaaaaa aagttttttt tcaactacag tatttttttaa
15001 aggcgcgcac ctttttgaat ttaacaaaaa tttgtcgtgt cgagaccggt tgccgtatat
15061 attgatttcc tgacaaaaat cacatgaaat ttctacggtg agcatggaaa aaaaaccatt
15121 tcaccacaaa aaatatagaa aagttgaaat tgactacagt aactttaaag gtacacacac
15181 taaaaccaac ttttgctgcg agacccaatc tcttaccaaa aagccgaatt aacagccatc
15241 aaatataatc cagtaagcgg attcctccat tcgatagttt ctcgatagac ttgatgaagt
15301 ctccggccac tttctcgtat ccgttccatc attttggagt ttaaattaaa attttaaaat
15361 ttcaacgaaa aaccaaaaaa agaaatcgat aactctcttc acatctgtta ttatttttctt
15421 ttttgtcaac ctaaaaacag aacagttgtt tctttttttc tttgaaactc ataagagata
15481 gaatgggagc gatgagaaaa cgcttcggga attattgaat tattgcgaaa tcggcatttt
15541 tttacaaatt tttttatgag atctggaatc tggaaaaaaa gagaaaaaag atgttaaata
15601 gaatctcgga tcaaaacata aaaaaataaa ataaataatt tattcccatc tgcttcgttt
15661 tgaggaagag ctgctgctcg aagagccacc agaagctgaa gaagcctgtt gctgtgcaat
15721 ctttgcattc aattcatcca ctttcttttt ccattctggt cttggatctg atgctacttg
15781 ctgtggaaga tttgaagatt catttgtagg tcgttgaaat gttgtttgga atgatttctg
15841 aaatgcactt ttgagcattt gggctctatt ggctccggaa gctgggccac tgcttgaaga
15901 ggaggaggaa gaggatgacg tggcaaagct gaaaattcga tttttatgtt aaatgtggtg
15961 gggtggatga aattattttt ttaaatactg tcatttatta gaaatttttt ttgttgaatt
16021 ttttggaaat tgttgaatta tttttttaat tcctattcta tttaaaactc aatttcccac
16081 atttttaaac atgctccgta taaaaaactt gataatttct ctgaaaaatt tcctaaaaat
16141 caataattta aactttgtcg cgtaaaactt ctgaaaattt gattgactat gtttgaaatc
16201 aattatcaaa caaaaaaatt taagagaacc aaatttttata gccgaaaacc cccaaaaatg
16261 tcgaaattag gccatttttt gcgggaattc aaagataaat cagcccgttt ttgtgggaaa
```

16321 attgaaaaaa aatcgtaaaa aatttagaat ttagattaac acttatattt ggctatttcg
16381 gtctattttc aatttttga agtgcaaata actctttttg aattaaaaaa tgattttctt
16441 tctttttcga aatatagttc aaaatttttg ccacttttt ctttttttta gtaaaaaatt
16501 tatttttttt cttcagcacc aacggtgtcg gtttttcgaa ttttgaaaac cgataatttt
16561 cgttttattt ctaatttccg atatttagta aaatttccg atttttaaaa aaaatcttaa
16621 cattttgatc tcttttctcc taaatttta tgtgtttttt tcttatcaat ataaaataaa
16681 attagatgaa aatttttaaa aatattttat ggttaattct ctgaattttt ttattttaat
16741 ttttcggaac atcaaatact caaaaattgt tagtttttcg atttcccttc aaaaaaatgg
16801 aaaaacgaca catttcttta attttcaaaa tataattttt cgattttccg gaaaatccgt
16861 aaactacaat ttttacattt tccttgtctc ggtagctatt ttttttatta aaaaattaat
16921 taaaaattta taccttttgtg catttctgat cattccatca actgtctgat tccctccacc
16981 tctacaactt ttttgctctt ttaatggatc aaattgtgct ccagtggcac ctcctccaac
17041 ttgtctaact tttggtgtat atccaagacc cattcttcca cgtgtctgtg ttccagtact
17101 tggtcctcca ttcccagctc tttgtcctct aaaccatgaa ctcttcattg caagatccat
17161 taaaggtttt ggtacttctt gtgatacact ttccagattt ttgaccaaat gtccgaccat
17221 ttcaatgtct ttttcagtaa caagagtgta tgctgttcct ttgtgaccag ctcttccagt
17281 tcttccaatt cgatgaacat gagtatcgat gtcacgagcc atatcgaaat tgattacggt
17341 tcgaatttca gaaatatcca gtccgcgtgc tgaaaattta atttttgat cggaaaatta
17401 gggaaaattg attaaaaacc ttagatttta gatttttaga actttaaaaa tttttttacg
17461 gtttccaaaa ttttcaaaaa cttgaaaatc gccaattttt aaatgttttt attggagaaa
17521 atatttgttt tttttttcct attttttcc gaaaaaatat attcaatttt ttaaaaaaaa
17581 tttcaataaa tcgagcaatt tttctcgact tttcgatgta aaaaaaaacg aaactttaaa
17641 agatttttag gaaatttttc aaaaagtcaa aatatttta tttttgttcc cgaatttcgt
17701 tttttttttc aatcagattt tttggttaaa aatctatatt tttttctttt tttccgaaaa
17761 aaaattaatt tttttggaga aatctcaata aatcgagcaa tttttctcga catttcaatg
17821 taaaaaaaaa acgaaaattt tagatttttt cataaatttg ctaaaatttt gaaaaaagtc
17881 aaaatatttc aatttttgtt gtcttaattt ttattaaaat ttaattgtgt taaaaaatcc
17941 cgaatttcat tttttttgta ttttttttctg attttcctgt taaaaatcta tactttcttt
18001 cgttccaaaa ataatttcaa ttaaaaaatt gggcgattct ttttcaattt tttcaattaa
18061 aaaaaattct caatatccgg ttctttttg atttttgtta acatttttta ttttagacaa
18121 ccgaaaatag gatttttgat tttacaaaaa aataaattaa ttcacttttt acagaaatat
18181 acttttttcgt tatttttttt cacatttttt aaattaaaat attatatttt taagattttt
18241 ggaattttcg aacttcaaaa aaaaattccc agtaaaacta ttacctgcca catcagttgc
18301 cacgagaatt tgcgattttt tccggaattt caataaattt tcatttctct ctgcttgtag
18361 catatctcca tgtagaagga caatgtcgaa atctttcatt ttcaactttt tcgcgacgtc
18421 ttccgaatcc aatttttttcg tgacaaaaat tagcactttt ccgactgaaa aatggtttaa
18481 attatggaaa atttgataat aataaaaatc ttacgagaag caaattccac taaatttcga
18541 ataagccaat gcaatttaac atcttgattt tgcattacaa aaacttttttg ttcaatatca
18601 gcattcgctt ctccaacttc tccttgtaca attcgtacgg gatctacgag agcatctcgg
18661 gcgagtcgtt ccacttttttg tttaaaagtc gcactgaaca ttaagcattg acgatccggg
18721 cggacgtgat cggatattga ttttacttga gcctctggaa taatggaaat tcgatttttt
18781 ttttcaaatt ttcatctttt tttttttgat aattgtatat ttaattgctg tgaatttagt
18841 aaccaaggaa ttccaaaaaa ttccaaaaaa aaaaacaagc tttgtggttt ttttttcaaa
18901 gaaaaatttg gctttaacgg tttttaaggt tttttttcaa aaattattca tttgaatttg
18961 ctacgtgtta agatgaaaaa tgttacggca acgaatccat aatcgagcca tttttctggt
19021 ttaaaaactc tcacaaaact caatgttttt ggttcgaaat ttttcgtcgc ggtcatgttt
19081 cgatttacga ggctcgtggt gggcaaacga ctttttccgg caaatcggca aattgccgga
19141 attgaatttt ccggcacttt gacgggattg aaaatttccg gcaaatcgac aattttccga
19201 aaatgaaaat ttccagcaaa tcggcaaact gccggaatcg aaaatttccg gcaaatcggc
19261 aaattgccgg aattgaaaat ttatggcaaa tcggcaaatt gccggaattg aagtttcggg
19321 aaaatcggta attttgttga aatggaaaaa ttccggcaaa tcggcaagtt gccggaattg
19381 aaaatttccg gcaaaccgac aaattgccgg aatttaaaag ttccggcaaa tcggaagatt
19441 gccggaattt gaaatttccg gcaaattacc ggaattaaaa tttcgggaaa gtcggcaatt
19501 ttgccgaaaa ggaaaatttc cggcaaatcg gcaattttaa atttgtcggt ttccactggc
19561 tcgtgtactc ctcgaggagg aaattcaatt cgggaaaata cacgagctgc gtaaatcgac
19621 atacggccat ttctataagt ttttttcaaa aaaatatacc aaatcccata tcaaacatgc
19681 gatcagcttc atcaaaaact aaaaaagttg tccgcaggaa attggtggct cccatcttca
19741 ctaaatcgat gattcgaccc tggaaaagcc ttttgaaatt taaaaataaa tatttttctc
19801 aactaaccgg tgtacaaaca accatttcag ctccttcatt ctgtaattca tttgattgtt
19861 cccatttact acctccacca tatgcacaaa ttggattaat attgtaaact ttgcagaatt

```
19921 ttttggcctc ttggaatacc tggaaaattt atttttttata tttcaagttt tttggttgaa
19981 tttccaaaaa cctgaattgc caattctctc gtcggaacca caataactgc aaccgggcct
20041 tctccagcct tcaaatcggg ttgatccata atatgaacga ttgcaggcca caaataggcc
20101 gcagtctttc ctgatccggt tttggcaatt cctagaacat ctcttccaga taatgcagag
20161 gggatcgcct gaaaatgtta atttttttgt cgataaatta cagatcgtcg atttgccgga
20221 aattttgatt ttcggaaaat tgtcggtttg ccggaaattt tcaattccgg caaggaaaca
20281 ttcataggat gcgtacaatt ttgccgatta aaattgaatt attccgtgac aatgtgcaaa
20341 cccacagttt gccgaaaatc gaaatttccg gcaaatcggc aaaaaaatct gaaaaatcaa
20401 tattcaccat ggcctgaatc ggagtcggct gctcatattc actcttccga attgcttcca
20461 ttaacagctt atcaaatgaa aaatgagcaa acgagcaaac tggacgtggc ggcttgaggc
20521 ctccaactcg aagattcatt gtattttgaa ggcgaattac atccatatag tgtaagcgtc
20581 tgaaaatttc ttgattgata atagaaattt tgctccaaaa ctacggtaac cggtctcgaa
20641 acgacttgtt aaatgcaaaa aagtgtgtgc gcctttaaag aatactgtaa tttcaaactt
20701 ctattgctgt ggaatttttt taatcgattt ttcatagttt atgaaaaatc gataaaaaat
20761 tccacagtaa ttaacgtcta cagtactcta taaaagcgca cgaacttttt ttgtaatcag
20821 caaaaatgtc gcgtcgagac ccaatttttt tcagaaaatg cacagaattg gcttatttca
20881 acatgataat cgaacttttt cgatttaaaa ttgatataaa aatgtagaaa gcgaggtttt
20941 ctatctgaaa ataatattta tttttgttga caaacttgaa aattaccgtt cgctgaaatt
21001 tttttaacgg caccatttga ggtatttttc aagattttca tcaaaaataa atgttatttt
21061 cagatagaaa accttgtttc ctacattttt atatcagttt tagttcgaaa aagttcgatt
21121 attctgttga aataagccaa ttttgtgaat ttgccgcgaa aaaaaatgtc ttatcctgaa
21181 cccacttaat atcctcatgc tcctcataaa agtttttatt aaacttttga tattgaatct
21241 gtgaatgatc gatatctggc aatggatcga taacctttt ccatgaccaa ataatatttc
21301 catcttcatc atattccaat tgctcatcat cattttcgtg tttctctttg tactcttcca
21361 tgaatctgga aaacttttat attgtttaat atttgatttt aaattcatga attaattcaa
21421 aatttcctgt acaaacttga atagagattc ctgcatatcc tcctcatcaa tatctgctct
21481 tccaagccct tttttactcg gatcctccgt atcttttccc tctttccgat ctttctcttt
21541 ttgctccgaa actttttttat cactggaagc ctgcttttcg attccggcca taaatgcgtc
21601 gagctcatct tcttcatcat cgtcatcggc ggcaagtttt gaggaggaag aagcgggctt
21661 ttcatcatct gagccgcctt ctaaatatct aaatttgaaa ttattaattt tctttttttt
21721 tttttgaaaa aaaaacctac tcatcatcat aaatccgttt ccttcttgat tgtttctcag
21781 acgtcgagcc ataagccata agatttgcgt actaaattaa taataatcca gtgaaaagtt
21841 gaaaattttt aatttttttca tgaaaactaa aaacctcttg ttccttcttc atttcttgct
21901 cttttcgatt tgacaagaga atattccgat ttggtggtgg aacaggcgca tttccgactg
21961 ttaaagctgc aggaggaggt acattattca tacttttacc atctgagaac aataaaattc
22021 atttaatacg tgaaaatatc gatttaccac tcaaactgga ggttcccgaa ctctttgcgc
22081 ttacaaatcg attttgatat cctcctgaat attttccacg ccacattttt gcctgaaaac
22141 ataataaatt ttgtttattt tggaggaaaa tgatgtttaa acatgtttaa aaagtcgttt
22201 aatcgcagtc acagggagaa tgtgggcgtg gccttcgctt gcaacacgcc gcgcgcattt
22261 cgcaacgctg ccgcacgtag tttatgtatt gctgaattgt tgaagtacgt aactacagta
22321 agaaaatcat tttttccaca ttgaaatatt tcaaaaaaaa aagagagaaa atattatatt
22381 ttgaaattat aagaaatgtg ggtagaaaac acaaaaacta tgaccttaag gtcgcacgag
22441 aaaattatta aaaataatta caacggaaaa actacgagaa aattacccctt tcaccatgaa
22501 aaaaaaacaa tttttatttt ctcccgcgac tttaagttca caaaagtttt aaaatattag
22561 cgtggcacac taatccatta gttaaactct aagctcttgc tatataaggg gcagaaatac
22621 ggttttcaaa tcagtacgaa gttttggaaa cttctaatct tctggcagtg gcagtgggga
22681 gcagttggcc cgtggatggc ctgtgggtgg cctgtggatg tcctgtggat gtcctgtgga
22741 tggtcagtgc atggtcggtg gtgggatgtt gccgaacaac cagcggcatc aatggtcgtg
22801 tcggaagccg aggcctcgga cgaaaaacac gtggatccgg atggaatgga caggatcggg
22861 tggtagcagc ggacagtgac caatattgaa acgtaagttc gaatgctact tttcagacat
22921 ttctgaaata actttaacaa gaacgaagac gtacgctaac tatatgttaa ctcgtctttt
22981 tgttttgaaa tttgaagtaa ttttgaatac aacagtggca aaagttaact gagtgagcaa
23041 ttattctttg ttttttgaaa ttaaaaagtg tgcataacag ttcctaaaca attcgaattt
23101 cagcgtggca cactgagaat gtgaggagtc atggagcacg ttgatgaggg tggagctgga
23161 acgatcggtt ggacactgga gtaagctgat aggctttaaa aggtaatttt tagaaatgaa
23221 acaattttga aaatagtaac tgcgagaaaa aaccatttca cgcagaattt aattatttaa
23281 aacacacatt agttcctaaa caataccagt tttcagagtg gcacaataaa cggatgggcg
23341 gtcaggcatg tcgtcgagca agcagcgaag cgataccact ttggactggg aaccgaggag
23401 catgtggatg aggctggaat ggccgggtag ctgcagagga cagtggtcaa gaatagaatg
23461 tgagtttatc agttattttt cagaaaaatg tttaataaga atgaagacga gagctatact
```

```
23521 tagctattgg ttaatttgtc gttttgtttt gaaatttgga gagcattttg aatacaacaa
23581 tgacaaaagt taattgagtt agcaattcgt ctttatgttt agaaattgaa aaagtttaaa
23641 taaaaaactg cgcgagaaag tgtacataat agttccaaaa caattagaat ttcaggatga
23701 cacaataaac tgatggatac tggatgagta gagggacctc gagaaacgta gcgactgaaa
23761 gatggggcgg gggcagtgga gaacgtggat caggatagat tggagcgatc gggacggtgt
23821 gcaaatggca atggagtcag tcaaggatgg actttagtcc aacaaaggta acgagggttt
23881 gaaatgattt tgaaaaaaaa gtaaaataaa atttttgatg aaaattaaag acactaataa
23941 aatattattt tttattcgtc ttttccaaat cagtgtgtcc atttttaaag tgtctactgg
24001 aagcttcatt ttttgaattc cataaattaa aatttcattt ccaaatcagt gtgtccactt
24061 ttaaagtgtc cactggaagc ttcatttttt gaactccaga aatttaagtt tcatttccaa
24121 atcagtgtcc cgtggtggaa aaatttttt ttccttttt ttgggatgaa aaaattcgca
24181 aaaaaaattt ttcatcccag aaaaaaaagg gggaaaaaag gttcagaaaa aacaaacaaa
24241 catttttgtc gctggatgga ctttagtcca acaaaggtaa cgagggtttg aaatgatttt
24301 gaaaaaaaag taaaataaaa tttttgatga aaattaaaga cactaataaa aatattattt
24361 ttattcgtct tttaatttta caaaatttc aataaaaatt taagacacta ataaaaatat
24421 tatttttatt cgtcttataa ttttagaaaa ttttcaataa aaatttaaga cactaataaa
24481 atattatttt tattcgtctt ttaattttag aaaattttca ataaaaatta aagacactaa
24541 taaaaatatg atttttattc gtcttataat tttagaaaat tttcaataaa aatttaagac
24601 actaataaaa tattattttt attcgtcttt taattttaga aaattttcaa taaaaattaa
24661 agacactaat aaaaatatga tttttattcg tcttttaatt ttagaaaatt ttcaataaaa
24721 attaaagaca ctaataaaat atgattttta tatatctttc tttctttctt taattcattt
24781 gtctttttt tcaattttga aatagttatt taatttggtc aattttgttt cttaaataat
24841 cttgataata aaacatcgat taaaaatatg attttatgtt tcaaaaaatg aattcaaatt
24901 gaagaattga gtgtggagaa gggggcaaaa ttgtaatgta caccgttgga gagaagagaa
24961 gccggctata tcacggacta tatcggctgg aagtagggga aaaggagatt actgtatttc
25021 ccctaataga catgcacctt taaagtgcac ttctattaga taaaagctcg atttttatga
25081 aaaattataa aaaatttcga aaaaaggggt ggttgaaaag gggcaatgac cagaatcgag
25141 atgagaactg attttgattt ttttttgtta aatttttggt attgaaaaca aacgggtgcg
25201 gggcggttag ggtaacagaa gatcacagaa tttggcagaa agaagacgca aaaagtagac
25261 aaagctagtg gagccgatcg gaaaattgac aagaaaacga tggaatgaag gaataaaaga
25321 aacacttgtc gtcggctgat atcccgtttg agaagagaga gagagaggaa gagagaggat
25381 ttcagaaaga aaggaaaaat tgggatgttt tctggaagtt tggtcttaaa gtagtttaac
25441 ttttaaagac aaaattccca gaaaaggcca gagaaagaaa aagaaagagg tagagaggca
25501 aaaaaaatca tcgaaaatca tcaataaaat aattgaatgt agatccgaag actatgagag
25561 cacacatcga cagcaacccg cggattttcg tttcgtcagt gtttctggaa aaaaattaat
25621 ttactttgaa attttcgcaa aaactcgaag tgtagctgga ctatttacta tgtatttact
25681 atggtggaaa aaaataggtg gccgagtttt ctttttttta cggccacgat ctttctcacg
25741 ggcggctatt ttttattaga acttttcatt tacaatttga gtcatattta tcgatcaggg
25801 aacatgcaaa aataatttaa aatttagaat ttggaatttt tttagtcaaa aaagtgtgtg
25861 tcttctgaga ttcttctgag cttcttcaag cttctgatat tttttttttc aaaaaatatt
25921 acaaagttat acttgttagg gtttaaggta tttaaagcaa ataatcaata aattttgaaa
25981 ctacagtact cttcaaaggc gcacaccgtt tgtttttatt gaacatttgt cgcgtgccga
26041 ggcctctgga ataaaaatca caaaatttcg caccgatgta atatattaat tttatgattt
26101 gttaaaattg aaattttaga actttcgcat ttgaaaaacc ttttatgtga ttttttaacta
26161 ttttaaattt aaaaaaatca gctcatgcga acgcgcttca gtaattagaa ataatcattt
26221 ttctgtcggt attttttaac tcaaattttt acaaaaaatg ttttacactc tccctaaatt
26281 tcaactttct agattctcaa gaaatctgcc aaaaagcatc acaagtactg agttcccatg
26341 ggaactgttg aattgtactc taaatatgtg tttttcccga gaatatggtg tattctttct
26401 tgtttttata agccccgagc actctaaaca tgtaaaaaat tccccgggga atcccatagg
26461 ctctggttac tccatctgtc agactaaact gtgttcacat taaattaggt caatttcctc
26521 tcaaaacacc tgcacaatcg aatcaaaact gaccacttga attaggttcc ggcaaatttt
26581 ccctagtaac caaatgcatg acggttgttt tgccgggcat tagctggagc gcgtgaagag
26641 ttacggaacc atgtagaaat cgaccgtgat agattaattt gagcatttga gcactttgta
26701 ctttatcttc ataccattct gaaaaataga attattcatt tgtgatagaa ttttttaaa
26761 aaattctgat cggccccgtc caaatcttgg cttcttttgt ttctatcgga tatttgtact
26821 gagaaataat ttgaaacaga aaccgagaac cagatggcaa ttttaaaaac ttaaaattca
26881 gagtcccgcc acgaaaacac cccgaaacac ctcagtttct gaaaacagga attaattttt
26941 cttcaaactg aaattcaaaa atttcgctaa tttcaaattc aaaaccagct aaaaattaat
27001 tttaaaaact gacaatatct acagattttc gagtaaacaa tcttttaggg tcccgccacg
27061 aaaacacctg aaggttctaa atatgttatt gtagactgaa aagcacatat ttcaggtggg
```

```
27121 aaattattaa agttcatcaa atttaagtac attttctgcc gaaaaaactg aaaaacttga
27181 aaaataaaat tttctgagaa gaaagttctt tcagagaata aagttagatt caaaaatttt
27241 atgaaaattt tttccaaatt ccattttgaa taaattgtac aatccctttg ggaaagtagg
27301 cgtaggcctc ctactgccta taattttaat gtttccctta ttcatttttt tttcaaattt
27361 ccataaaccg aagttgcaac tgctacaaca tgtctcgttt cgcacatatt tgcagacaca
27421 tacacctgac ctcctaagat cagtggaatt gtggtgttcc cccacaacct tctcaagacc
27481 cctctataag cctgtagcaa acaatagtct atgtttctcg gacatcgtct ttgtgtactg
27541 gcgacaggaa taatgagccc gtgtgtgcgc gcgcgagcgt cgaacatttc ggattaacac
27601 acactcactc acaaaagttt caaagttact taccatcagg ccattgatcg aataccattt
27661 gtgtaacgtc gtgagctgat gtgagtgggt ggaattcgaa ttcatgagtt tttccgctga
27721 caaggatgag acgaagtacc acctaaaagt tggatgtgtt atagataggt ttaggcttag
27781 gttcaggctt tggcttaggc ttattcttaa gcttagggtt aggcttaggc ttaaaaacgg
27841 gaaccgttca gatttatagg ccatgtgaaa aaatgtttaa actaatattt tttcaaatta
27901 actcttaaaa attcaaattt taatttccg ggtaaatttc ctttttttatt agtggccgaa
27961 gctgcagttt tctaggcggc cgccaatttt tttaggccat gtagaacacc actagttata
28021 attgttttga aacgcggcca ccgaaatttc taaaatttct aggccatgta ttacgtcaaa
28081 ttttctctag attttttttca aacaaaaaaa actcactcgt tctgcttgag aatcaatcga
28141 ctgttttgca ctcatatccg ggtgcttttc tgattaagaa aaaattcttc tcctgatata
28201 tataatttgt ttcaacgact tcagagtaat tcacgtcgtt tccccacaca cacacggtgt
28261 caccctgaaa aataacaaaa ctatgagcaa atcgatgaga gaaacaggac gtgcgaggac
28321 ggaatgaggg agagaaaagg gactaattat attcatagat gacttggaat ggaattgaaa
28381 caaagtacag agtattctta ttaactggaa agttagatga taaaaaagag aaaacgacgc
28441 aaaaaataac gagggctctt ctgaatgcaa aaatgtaaat gaggttgaga atactgatga
28501 gagagtgatt cttgtccaag ttttgtacgt ggccgagaaa agaagaaaat ctcggccacg
28561 cagataaatg gggtttacag ataaattagt tgaagcttca aaggtgaaag ttttaaggaa
28621 gatagaaaaa atctaagtgg cctagaaaac tctttcacgg ccatcaactt caaatatgtc
28681 tccagccact ccggcagttt ttttttaata tgccgccatt tttatggacc tctatgaaaa
28741 aaaaaatcaa tttctcaaat tttaaaaatc taaaagctgt tggaaaaaaa tgaaaagaaa
28801 aaagaaaatg aaaccaaatg ttagttagtt gagaaacttg aaaaaagaaa cgagaataag
28861 agactgagta aatattgcta agcgacagat ttcgaacaga tgaaatgaga aaggggtct
28921 agtggggtag aagcatagag tatgagagtg aaggagagtg agagagcacg cagagaaaaa
28981 gagattagtt ctacttttca gacgtgcgta cgttcttagt catgcagaat tctgagaata
29041 ttcaaaaata aggaaatgag taataagaca ttagattaag taatagtaga gttcagaaaa
29101 ttgagataaa gctgaaattt taaaaagggt ttaaggccac ctttaaggta aatttccgtt
29161 gtgaagtttt ctaggccacc aacataacca acggctcaaa atgtttcaga acagctcttg
29221 tttttctaga cttttagact tctctttttt aaattgcaca aaaaattggt cttttcgtga
29281 tcatcgagac cacttgaaag tacattgaaa agaacaacgt tgcggcgtca ctgagttatt
29341 tttgaagttt tgttgggcaa cgaggacggg ggaagagaag aaaacataaa atcctaaccc
29401 taacccgacc ccaagcctaa gcccaagtct aagcttaagc ctagcccaaa cctaagccta
29461 aacctaacat ttaattaatt ttcaataatg acagctcaaa aaaaagagca aaaaacgggc
29521 ggacctaact tgttcgacta aagggcacgc caagctcaaa gagcattaaa tgcacgttca
29581 gactacttgg acactcaggt caggttacgc gtggaaatac tagaagctct ttttgatgac
29641 gtggcatttt ttttgtctga aaatggagag ggagtggcgg acgagggaga aaaactttgt
29701 atgcataaga aaaactgctg ctgctacttc agcttctagc aaactctaga ataacatgtt
29761 gacgatatat atttttcccc gttttctgaa ttatttgaag aggcaaaaat attataatat
29821 tgaatgaaaa actaggaaat tttagaattt gagttagttg caattgcgct ttaattatat
29881 ttcaagcctt tctgaagcct gtggaagagt gctggggcat aattttcatg gcctaacttt
29941 cctgaagcct cggtcgtcaa aatatcgaaa aaaaaagtta tgggttttta aaacgaaaaa
30001 tattcaaatt ttttcagttt aagaaaatcg aaaacaacaa aaatgtttct aaaaaattct
30061 aaaaatttta ttcttattta ttgattcaaa aattgaaaat tttggatttt tttttctggc
30121 tcactatcag atttcttgat tttccttttt ttttctagaa aatttctatt tttttaaata
30181 aaaaaatgtg gaaatgttat ttcttttgat acataatata catggtacga tacattttaa
30241 atattccaaa tcggccaaat atcatgataa aacactagaa aatttttaaag gttttttggc
30301 aaatttctaa agcctacaac tcctagaata tttaacaaaa aatattttaa acgcaattaa
30361 aatataaaaa tcgtagagaa atttttttgtt ggtcgagttc caaaaaataa aaaaaattgg
30421 aaatttggaa gaaaaaaaca tttttttaat ggttctttaa tcaattaaac tttaattatt
30481 ggtatttaaa tattttgagg aggtggacat gcaaatccta aacaaaagca aaatgcgttg
30541 tcaatttaaa ttttagtttt ttttaaatat gaaaaaaaca caaaaatttc agatctaaaa
30601 tagaagttgt caaaaaattc tcaaaaaaag tgctgaaagt gatggctgac ctcgaaaagg
30661 gaaatgttat gggtataaga gagagagaga gagagtggcc gcaggaaaaa aaacgggttg
```

```
30721 gcgccccccg gagaagacga gatgcgggac gaataatttg gttttgacaa cgacggaaaa
30781 tacgaagaag acgaagaaga agacgacgat gtgtatgttc ttcttttgtc gaaatccggc
30841 ccttttggtg gtgtctgcgt ctctctttct ctctatagaa atacactcca agcacacttg
30901 aggtaaataa aattgagaag tggttgtggt ggtggcacat aggtgtttgt atgatgagat
30961 tgagagagag agagtggatt aatataattg attgattgaa aatgtgatcg agagggaaat
31021 tcgatcgaga gggatcgaga gggattgata tggcgaccct agtgatgaat aaggttcagg
31081 aaaactagca taacttttta aaaatgtcat ctaacaattt ttttaatatt ttttctgta
31141 agaaaaacat tgaaaaatac atagttttgg tcacttttca gttgtattgg ctaagaatca
31201 agtaagatat ttcaatttaa attcgaaaat tgcagtaggt caactctccc cgaccagaaa
31261 tggttaggtt tgtggaaagt gacataactt tgtagatgcg tcatctaaaa aatttttaa
31321 tgtattttat aaaagaaaaa catttaacga tattgttggt aagttttcaa ataaattggt
31381 taaaaattga gtgagacatt ctagtttaaa ttcgagtaat atccgctttc cactgcaaaa
31441 gattaactaa ttgaattttc caatattttc taagattatg aaataatttg taaaatatct
31501 taataaccac attttgtaaa tatgaaataa aaaagtttca acaaaaaaat cgataaaaac
31561 tttttttttc gaatttttcg aaaatttaaa ctctaacaat aaactcgggt atttttagc
31621 tttaaaaatt atttttgga atcaaatcta cgaaattatt caaattagcg taaaaattct
31681 gctcaccaat cagcggcgtt gcccacgccc ctgggcaacg ctttaactct ttttactcat
31741 tttggcactt ttttgagaca ttttttgaag attggcttta cgtaaagtaa tttcacatca
31801 acatttttg tattgcctac tgccaaatat tgtaggcacg taggcaggca cacatgccgg
31861 caaataggca tgctccaatt aaagatcact cctctccaaa tattatgggt attggtgtct
31921 agagagcaac aacaaaaaat gggaagcata aagcaattca aagattcctc ctttttttct
31981 tttctcgcac gcgatagatg tttggcaaaa gaaatgacgc gtggctttcg atatctgaaa
32041 aaataatatt caaaaaaaaa aaatattttt gttgaaaagc tgatgagaaa aagaagaaaa
32101 ggaaataaaa ggtgataata ggaaaaagag gaacaaatga atgatgtgaa gaccggagaa
32161 ggaggcgggt tatcaggtta ttatcacata caaataagcc atattatcag tcaggaggag
32221 cggtatatgg gaacattgct atttgggaaa aagataatag ggcaattatt caaattatga
32281 ataggggaac aaaaggtttg aaccgaccat tgagatgggg ttggagaaca ttttttcaat
32341 tttagcggaa aattcaaatt ttaattacaa aaacacatac ctacttagag ggaaattcaa
32401 ctttttttaat tttttgaaaa ttaatgacaa aattgcatgg gtttttgttt ttcaaaatat
32461 agtattacgg aactatgaga tcatgagaat acctactctt tgccacatgg cctctcgctg
32521 ccgcggcctg ccaatatttt cgcgccagtg aataggcatt ctcatgatct catggtcccg
32581 taatatgtta ttctgttttt aatatttcaa aaaattttca caggaatttt tttattcaaa
32641 ttttaatttt tttttaattt acaattaggt agaaaactta aattttaaga tcaaagtttt
32701 ggcgggtaac ttgaattttt aaaaaatatt ttgcagaaaa cacaaattaa agcttattaa
32761 tttaaaggga aattcaattt ttaattttca aaaacattcg aaattagttc aattttcaat
32821 ttattttcga atttttaaa aactatcttt ggcgggaagt tcaaatttca attataaaat
32881 tttgccagaa atttttctca ttttttgttt ttaaaaaaat tggcgttaaa aacaagatga
32941 gttaaattca atttttgatt aaaaataatt ttggtgacag gaaatgaaaa tttaatttgt
33001 attattatgg caggaaatta aaattttaaa ttaaagaaat attctgaatt aaattcaatt
33061 ttttaagatt tgttttgaat agctttttgc caaaaaaaat gcaggaaaaa atgcaggaaa
33121 ctcaaactga aagtttttc aagtaaaatt ttaaaaaatc aaaattttgg aaggagctat
33181 tatttttat tttgacggaa aaatcaaatg tttatatttc aaaaaaattt ggtgtgaaat
33241 tcaaacttgt tttaaaaaat tccaaaataa tatttttagc agtaaaaatt ttaaagatgg
33301 ataaaagtcg tatgatcata aagtttccta ttataacttt ataaaaattc tcttccgacg
33361 gaaaatagaa tattcgtaat tctaaagtct atttggatat agtgaatggc tgaattcggg
33421 tggcttcatt gacggtgcac tgaagcgaga aacgacttga tttggttggc attcaacctt
33481 ttcttttctc tttctaatgc aggtcaatca gtgaagacag agagagatag gcgtatagga
33541 aagaagaaag aagaaggagg aaaagggaat tggcaaagaa aaggcataac aataacaaca
33601 gaaagattca acgtcgtcac aaattgaatt gggaacgaac aacaccaaga aacaatcccg
33661 tacacttttt ccagaaatgg aacttttttac agaaacattg aaaaaaaaac cacatgattt
33721 caagaaaaac cgcaagtaat ttaaaaagga accaaatttt ttcagctgct caaaaaatca
33781 gctggaaata atcaaaaatt cactttctgg ttcctttttc gcctcccctt tccttctcaa
33841 atgaatcata ttgcaaagta aaagtgagac agaaacattg aattatttat gaaaaaattt
33901 ggtgggtttc atgtgacttt tagattttaa agtaattaaa aattaacatt ttcgttttta
33961 caattccttg atttagttgg ttttggttta cattaaaaat tgttaactaa taaattacag
34021 attatgtttt ttgctacact ttcgtagttg gacaaagtta aaagggatat attagtattg
34081 caccccgggg gcctcaactt tggaaactta tatctcagtt gttttcaatt atattaaaac
34141 catgttaact acgtgttaat atgtattcta tcgttttcta tatttttgta tttagttgga
34201 catttttga tatcgactag aacaatttag ttaaacactg tcgtcaaagt tgatcctaac
34261 ttcaacttt tggaaaaaaa gtctaactgt atttttaaa caatttttc tttccttaca
```

```
    34321 aaaaagtaat aaaatgaaat tttcagaaaa atatttgaaa ttctcgaaag cagactaaat
    34381 tttgcagaac tttacaacaa atttcgcgga aaaaacccag agacctagca ctttctaaca
    34441 gtgaaaaaat caattcaaga aaataaaatg cttcttttca aaaagtacac aaaactgtac
    34501 gacaccaaat gaggcaagaa tgaaagcatt gctgtccgcc tgcgacaaaa aagtttcaac
    34561 gttcttttct gtcttggtca gggaacgggc ttacatcggt tgacaatgcg attgttccga
    34621 caaaagagaa aaagaagaag aaaaaacatt gaaatggagg ataagttttg gggtttttgt
    34681 tgggaggaat caatttgctt cgaaatctca aagtttcggg aaatttctaa tatttttaa
    34741 tgcaaaattt tggctgaaat tcgggggttt gagggttttt tttaatgaaa aaaatagcat
    34801 ttgaaaaact gaagttcatg agcttccgga tgacgttaaa a
//
```

22

1: X81823. D.discoideum Hel2...[gi:550328]

Links

```
LOCUS       DDHEL2A                 1748 bp    mRNA    linear   INV 02-AUG-1995
DEFINITION  D.discoideum Hel2A mRNA for RNA helicase.
ACCESSION   X81823
VERSION     X81823.1  GI:550328
KEYWORDS    DEAD box protein; Hel2A gene; RNA helicase.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1
  AUTHORS   Mahal,B. and Nellen,W.
  TITLE     Developmental regulation of DEAD box proteins and cloning of
            putative RNA helicase genes from Dictyostelium discoideum
  JOURNAL   Biol. Chem. Hoppe-Seyler 375 (11), 759-763 (1994)
  MEDLINE   95209781
   PUBMED   7695838
REFERENCE   2  (bases 1 to 1748)
  AUTHORS   Mahal,B.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-1994) B. Mahal, Max-Plank Inst.f.Biochemie, Abt.
            Zellbiologie, Am Klopferspitz, 85152 Martinsried b. Munchen, FRG
FEATURES             Location/Qualifiers
     source          1..1748
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX2"
                     /sub_strain="214"
                     /db_xref="taxon:44689"
     gene            1..1748
                     /gene="Hel2A"
     CDS             <1..>1748
                     /gene="Hel2A"
                     /codon_start=3
                     /product="putative RNA helicase"
                     /protein_id="CAA57417.1"
                     /db_xref="GI:550329"
                     /db_xref="GOA:Q23909"
                     /db_xref="TrEMBL:Q23909"
                     /translation="IPDFRSELGVKITGKDCPKPIQSWAQAGLTEKVHLLLKKFQYEK
                     PTSIQAQTIPAIMNGRDLIGIARTGSGKTLAFLLPMFRHILAQPKSAPGEGMIALIMS
                     PTRELALQIHVECKKFSKVLGLRTACVYGGASISEQIAELKRGADIVVCTPGRMIDIL
                     CANNRRITNLRRVTFLVLDEADRMFDMGFGPQINCIVDSIRPDRQTIMFSATFPPKVE
                     NVAKKILNKPLEIIAGGRSIVSSDIEQFVEVRPTETRFRRLIELLSIWYHKGQILIFT
                     NRQETTDNLYRQLSNSQYQCLSLHGSKDQTDRDETISDFKNKVKTILIATPLASRGLD
                     IKDLNLVVNFDCPDHLEDYVHRVGRTGRAGNRGTAYTFITPDEERFSSSIIKALEQSG
                     SKVPDELRKLNDTYEKKRKEGKDVLLAPTGFTGRGHKFDAAEEDKKNIERKQQRKAYG
                     IEEEEEEEDEDKEKAEKEKLAAASAEKEKQLLSEKEKLDPATTNTIVIPGVDGTIITP
                     SSLLQTDPSVPVGQQAINQIFGISQVTSSEEAIKKLQLAAQLGMKGNIQKLNNQITPL
                     NQTHFIEELEINDSEF"
     misc_feature    537..548
                     /gene="Hel2A"
                     /note="DEAD-box"
ORIGIN
```

```
   1 gaattccgga ttttcgatca gaattgggtg tgaaaatcac tggtaaagat tgtccaaaac
  61 caattcaatc atgggcacaa gctgggttaa ctgaaaaggt tcatttacta ttaaagaaat
 121 ttcaatatga aaaaccaaca tcaattcaag cacaaactat acctgcaatt atgaatggtc
 181 gtgatttaat tggtattgcc agaactggtt ccggtaaaac attggcattc cttttaccaa
 241 tgtttcgtca tatactggca caaccaaaat ctgcacctgg tgaaggtatg attgcattga
 301 ttatgtcacc aactagagaa ttggcacttc aaattcatgt tgaatgtaaa aagttttcaa
 361 aagtacttgg attacgtacc gcttgtgttt atggtggtgc aagtataagt gaacaaatag
 421 ccgaattgaa aagaggtgcc gatattgtcg tttgcacacc aggtcgtatg atcgatattt
 481 tatgtgcaaa taatcgacgt atcaccaacc taagacgtgt aacattcttg gtgttggatg
 541 aagccgatcg tatgtttgat atgggttttg gtccacaaat taattgtatc gtcgatagta
 601 ttagacccga tcgtcaaacc attatgttct ctgcaacttt tcctccaaaa gttgagaatg
 661 tcgcaaagaa gatcctaaac aaaccattgg aaatcattgc tggtggtaga agtatagttt
 721 catcagatat tgaacaattt gtagaggtac gtccaactga aactagattt agacgtttaa
 781 tagaattgct atcgatttgg tatcataaag gtcagatttt aatctttacc aatcgtcaag
 841 agaccaccga caatctatat cgtcaacttt caaactctca atatcaatgt ctatcattac
 901 atggtagtaa agatcaaacc gatcgtgatg aaaccattag tgactttaaa aataaggtta
 961 aaaccatttt aatcgctaca ccattggcat cacgtggttt ggatatcaaa gatttaaatc
1021 ttgtggttaa tttcgattgc cctgatcatt tggaagatta tgttcatagg gtaggtagaa
1081 ctggtagagc aggaaatcgt ggtactgctt atacatttat cacacccgac gaagagagat
1141 tctcttcgtc aatcattaaa gctttggaac aatctggttc aaaagtaccc gatgaactta
1201 gaaaattgaa tgatacctac gagaaaaaga gaaaagaagg taaggatgta ctattggcac
1261 caaccggttt cactggtaga ggtcataaat ttgatgctgc cgaagaggat aaaaagaata
1321 ttgaaagaaa acaacaaaga aaagcatatg gtatcgaaga ggaagaagaa gaagaggatg
1381 aagataaaga aaaagctgaa aaggagaaat tggccgctgc ttccgctgaa aaagaaaaac
1441 aattattatc tgaaaaagaa aaattggatc ctgctaccac taatactatc gtcatacctg
1501 gtgtagatgg tacaatcatt acaccttctt cattacttca aaccgatcct tcagttcctg
1561 tgggtcaaca ggctatcaat caaatatttg gtatttcaca agttacctcc tccgaagaag
1621 caattaaaaa acttcaattg gccgctcaat taggtatgaa aggtaatatt caaaaattaa
1681 ataatcaaat aactccatta aatcaaactc atttcattga agaattagaa attaatgatt
1741 cggaattc
```

//

1: P09052[gi:1352826] This record was replaced or removed. See revision history for details.

```
LOCUS       P09052                  661 aa            linear   INV 01-FEB-1996
DEFINITION  VASA PROTEIN.
ACCESSION   P09052
VERSION     P09052  GI:1352826
DBSOURCE    swissprot: locus VASA_DROME, accession P09052;
            class: standard.
            created: Nov 1, 1988.
            sequence updated: Feb 1, 1996.
            annotation updated: Feb 1, 1996.
            xrefs: gi: 433675, gi: 1054723, gi: 8804, gi: 84895, gi: 84894
            xrefs (non-sequence databases): FLYBASEFBgn0003970, PFAMPF00270,
            PFAMPF00271, PROSITEPS00039
KEYWORDS    Developmental protein; ATP-binding; Repeat; Helicase.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (residues 1 to 661)
  AUTHORS   Lasko,P.F. and Ashburner,M.
  TITLE     The product of the Drosophila gene vasa is very similar to
            eukaryotic initiation factor-4A
  JOURNAL   Nature 335 (6191), 611-617 (1988)
  MEDLINE   89014721
  REMARK    SEQUENCE FROM N.A.
REFERENCE   2  (residues 1 to 661)
  AUTHORS   LASKO,P.F.
  TITLE     Direct Submission
  JOURNAL   Submitted (~DEC-1993)
  REMARK    REVISIONS.
REFERENCE   3  (residues 1 to 661)
  AUTHORS   Hay,B., Jan,L.Y. and Jan,Y.N.
  TITLE     A protein component of Drosophila polar granules is encoded by vasa
            and has extensive sequence similarity to ATP-dependent helicases
  JOURNAL   Cell 55 (4), 577-587 (1988)
  MEDLINE   89028669
  REMARK    SEQUENCE FROM N.A.
COMMENT     [WARNING] On Feb 1, 2001 this sequence was replaced by a newer
            version gi:12644110.
            On Jun 4, 1996 this sequence version replaced gi:137452.
            ---------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------.
            [FUNCTION] THE VASA PROTEIN IS REQUIRED ONLY IN THE FEMALE GERM
            LINE. IT IS IMPORTANT FOR OOCYTE FORMATION AND IN THE SPECIFICATION
            OF THE POSTERIOR STRUCTURES OF THE EMBRYO.
            [DEVELOPMENTAL STAGE] MATERNALLY EXPRESSED (DURING OOGENESIS).
            FUNCTION DURING EARLY EMBRYOGENESIS.
```

```
          [SIMILARITY] TO OTHER 'DEAD' BOX FAMILY HELICASES.
FEATURES             Location/Qualifiers
     source          1..661
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
     gene            1..661
                     /gene="VAS"
     Protein         1..661
                     /gene="VAS"
                     /product="VASA PROTEIN"
     Region          35
                     /gene="VAS"
                     /region_name="Conflict"
                     /note="A -> R (IN REF. 3)."
     Region          93..127
                     /gene="VAS"
                     /region_name="Domain"
                     /note="5 X 7 AA TANDEM REPEATS OF [FS]-R-G-G- [EQ]-G-G."
     Region          93..99
                     /gene="VAS"
                     /region_name="Repetitive region"
                     /note="1."
     Region          100..106
                     /gene="VAS"
                     /region_name="Repetitive region"
                     /note="2."
     Region          107..113
                     /gene="VAS"
                     /region_name="Repetitive region"
                     /note="3."
     Region          114..120
                     /gene="VAS"
                     /region_name="Repetitive region"
                     /note="4."
     Region          121..127
                     /gene="VAS"
                     /region_name="Repetitive region"
                     /note="5."
     Region          265
                     /gene="VAS"
                     /region_name="Conflict"
                     /note="F -> Y (IN REF. 3)."
     Site            289..296
                     /gene="VAS"
                     /site_type="np-binding"
                     /note="ATP (BY SIMILARITY)."
     Site            399..402
                     /gene="VAS"
                     /site_type="unclassified"
                     /note="DEAD BOX."
     Region          582
                     /gene="VAS"
                     /region_name="Conflict"
                     /note="C -> R (IN REF. 3)."
     Region          594
                     /gene="VAS"
                     /region_name="Conflict"
                     /note="D -> H (IN REF. 3)."
     Region          644
                     /gene="VAS"
                     /region_name="Conflict"
                     /note="R -> RQ (IN REF. 3)."
ORIGIN
```

```
  1 msddwddepi vdtrgarggd wsddedtaks fsgeaegdgv ggsggegggy qggnrdvfgr
 61 igggrgggag gyrggnrdgg gfhggrrege rdfrggeggf rggqggsrgg qggsrggqgg
121 frggeggfrg rlyenedgde rrgrldreer ggerrgrldr eerggerger gdggfarrrr
181 neddinnnnn iaedverkre fyippepsnd aieifssgia sgihfskynn ipvkvtgsdv
241 pqpiqhftsa dlrdiiidnv nksgfkiptp iqkcsipvis sgrdlmacaq tgsgktaafl
301 lpilsklled phelelgrpq vvivsptrel aiqifneark fafesylkig ivyggtsfrh
361 qnecitrgch vviatpgrll dfvdrtfitf edtrfvvlde adrmldmgfs edmrrimthv
421 tmrpehqtlm fsatfpeeiq rmageflkny vsvaigivgg acsdvkqtiy evnkyakrsk
481 lieilseqad gtivfvetkr gadflasfls ekefpttsih gdrlqsqreq alrdfkngsm
541 kvliatsvas rgldiknikh vinydmpski ddyvhrigrt gcvgnngrat sffdpekdra
601 iaadlvkile gsgqtvpdfl rtcgaggdgg ysnqnfggvd vrgrgnyvgd atnveeeeqw
661 d
```
//

24

1: Q64060. DEAD-box protein ...[gi:2500526] BLink, Domains, Links

```
LOCUS       Q64060                    713 aa            linear   ROD 01-NOV-1997
DEFINITION  DEAD BOX PROTEIN 4 (VASA HOMOLOG) (RVLG).
ACCESSION   Q64060
VERSION     Q64060  GI:2500526
OBSOURCE    swissprot: locus DDX4_RAT, accession Q64060;
            class: standard.
            created: Nov 1, 1997.
            sequence updated: Nov 1, 1997.
            annotation updated: Nov 1, 1997.
            xrefs: gi: 806463, gi: 806464
            xrefs (non-sequence databases): PFAMPF00270, PFAMPF00271,
            PROSITEPS00039
KEYWORDS    ATP-binding; Helicase; RNA-binding.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE   1  (residues 1 to 713)
  AUTHORS   Komiya,T. and Tanigawa,Y.
  TITLE     Cloning of a gene of the DEAD box protein family which is
            specifically expressed in germ cells in rats
  JOURNAL   Biochem. Biophys. Res. Commun. 207 (1), 405-410 (1995)
  MEDLINE   95160706
  REMARK    SEQUENCE FROM N.A.
            STRAIN=WISTAR; TISSUE=TESTIS
COMMENT     ---------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------------.
            [TISSUE SPECIFICITY] TESTIS.
            [SIMILARITY] TO OTHER 'DEAD' BOX FAMILY HELICASES. HIGHEST TO
            DROSOPHILA VASA.
FEATURES             Location/Qualifiers
     source          1..713
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     gene            1..713
                     /gene="DDX4"
     Protein         1..713
                     /gene="DDX4"
                     /product="DEAD BOX PROTEIN 4"
     Site            317..324
                     /gene="DDX4"
                     /site_type="np-binding"
                     /note="ATP (POTENTIAL)."
     Site            431..434
                     /gene="DDX4"
                     /site_type="unclassified"
                     /note="DEAD BOX."
```

ORIGIN

```
        1 mgdedweaei lkphvssyvp vfekdkyssg angdtfnrts asssemedgp sgrdhfmrsg
       61 fssgrnlgnr digesskret tsttggfgrg kgfgnrgfln nkfeegdssg fwkestndce
      121 dtqtrsrgfs krggypdgnd seasgpfrrg grdseydqdq gsqrggglfg srkpaasdsg
      181 sgdtfqsrsg nargaykgln eevvtgsgkn swkseaegge ssdiqgpkvt yipppppede
      241 dsifahyqtg infdkydtil vevsghdapp ailtfeeanl cqtlnnniak agytkltpvq
      301 kysipivlag rdlmacaqtg sgktaafllp ilahmmrdgi tasrfkelqe peciivaptr
      361 elinqiylea rkfsfgtcvr avviyggtqf ghsirqivqg cnilcatpgr lmdiigkeki
      421 glkqvkylvl deadrmldmg fgpemkklis cpgmpskeqr qtllfsatfp eeiqrlagef
      481 lksnylfvav gqvggacrdv qqsilqvgpv fkkrklveil rnigderpmv fvetkkkadf
      541 iatflcqeki sttsihgdre qrereqalgd frcgkcpvlv atsvaargld ienvqhvinf
      601 nlpstideyv hrigrtgrcg ntgraisffd tesdnhlaqp lvkvlsdaqq dvpawleeia
      661 fssyappsfs nstrgavfas fdtrknfqgk ntlntagiss aqapnpvdde swd
//
```

25

1: Q61496[gi:2500525] This record was replaced or removed. See revision history for details.

```
LOCUS       Q61496                  637 aa            linear   ROD 01-NOV-1997
DEFINITION  DEAD BOX PROTEIN 4 (VASA HOMOLOG) (MVH).
ACCESSION   Q61496
VERSION     Q61496  GI:2500525
DBSOURCE    swissprot: locus DDX4_MOUSE, accession Q61496;
            class: standard.
            created: Nov 1, 1997.
            sequence updated: Nov 1, 1997.
            annotation updated: Nov 1, 1997.
            xrefs: gi: 286074, gi: 286075
            xrefs (non-sequence databases): MGI102670, PFAMPF00270,
            PFAMPF00271, PROSITEPS00039
KEYWORDS    ATP-binding; Helicase; RNA-binding.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (residues 1 to 637)
  AUTHORS   Fujiwara,Y., Komiya,T., Kawabata,H., Sato,M., Fujimoto,H.,
            Furusawa,M. and Noce,T.
  TITLE     Isolation of a DEAD-family protein gene that encodes a murine
            homolog of Drosophila vasa and its specific expression in germ cell
            lineage
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (25), 12258-12262 (1994)
  MEDLINE   95083681
  REMARK    SEQUENCE FROM N.A.
            STRAIN=BALB/C; TISSUE=TESTIS
COMMENT     ---------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------------.
            [SUBCELLULAR LOCATION] CYTOPLASMIC AND PERINUCLEAR.
            [TISSUE SPECIFICITY] TESTIS.
            [DEVELOPMENTAL STAGE] EXPRESSED IN SPERMATOGENIC CELLS FROM THE
            SPERMATOCYTE STAGE TO THE ROUND SPERMATID STAGE.
            [SIMILARITY] TO OTHER 'DEAD' BOX FAMILY HELICASES. HIGHEST TO
            DROSOPHILA VASA.
FEATURES             Location/Qualifiers
     source          1..637
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            1..637
                     /gene="DDX4"
                     /note="synonym: MVH"
     Protein         <1..637
                     /gene="DDX4"
                     /product="DEAD BOX PROTEIN 4"
     Site            246..253
                     /gene="DDX4"
```

```
                     /site_type="np-binding"
                     /note="ATP (BY SIMILARITY)."
     Site            360..363
                     /gene="DDX4"
                     /site_type="unclassified"
                     /note="DEAD BOX."
ORIGIN
        1 fgrgkgfgnr gflnnkfeeg dssgfwkesn ndcednqtrs rgfskrggcq dgndseasgp
       61 frrggrgsfr gcrggfglgr pnsesdqdqg tqcgggflvl gkpaasdsgn gdtyqsrsgs
      121 grggykglne evvtgsgkns wksetegges sdsqgpkvty ipppppeded sifahyqtgi
      181 nfdkydtilv evsghdappa iltfeeanlc qtlnnnirka gytkltpvqk ytipivlagr
      241 dlmacaqtgs gktaafllpi lahmmrdgit asrfkelqep eciivaptre linqiylear
      301 kfsfgtcvis vviyggtqfg hsvrqivqgc nilcatpgrl mdiigkekig lkqvkylvld
      361 eadsmldmgf apeikklisc pgmpskeqhq tllfsatfpe eiqrlagdfl ksnylfvavg
      421 qvggacrdvq qtilqvgqyq keksllrfye nigdertmvf vetkkkadfi atflcqekis
      481 stsihgdreq rereqalgdf rcgkcpvlva tsvaargldi envqhvinfd lpstideyvh
      541 rigrtgrcgn tgraisffdt dsdnhlaqpl vkvlsdaqqd vpawleeiaf styvppsfss
      601 strggavfas vdtrknyqgk ahveysgdff ftssqss
//
```

(26)

1: Q62167. DEAD-box protein ...[gi:2500528]


```
LOCUS       Q62167                    662 aa            linear   ROD 15-JUL-1999
DEFINITION  DEAD BOX PROTEIN 3 (DEAD-BOX RNA HELICASE DEAD3) (MDEAD3)
            (EMBRYONIC RNA HELICASE) (D1PAS1 RELATED SEQUENCE 2).
ACCESSION   Q62167
VERSION     Q62167  GI:2500528
DBSOURCE    swissprot: locus DDX3_MOUSE, accession Q62167;
            class: standard.
            extra accessions:O09060,O09143,created: Nov 1, 1997.
            sequence updated: Nov 1, 1997.
            annotation updated: Jul 15, 1999.
            xrefs: gi: 1835121, gi: 1835122, gi: 407995, gi: 407996
            xrefs (non-sequence databases): MGI103064, PFAMPF00270,
            PFAMPF00271, PROSITEPS00039
KEYWORDS    Helicase; ATP-binding; RNA-binding; DNA-binding.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (residues 1 to 662)
  AUTHORS   Sowden,J., Putt,W., Morrison,K., Beddington,R. and Edwards,Y.
  TITLE     The embryonic RNA helicase gene (ERH): a new member of the DEAD box
            family of RNA helicases
  JOURNAL   Biochem. J. 308 (Pt 3), 839-846 (1995)
  MEDLINE   97104282
  REMARK    SEQUENCE FROM N.A.
            STRAIN=C57BL/6, AND DBA
REFERENCE   2  (residues 1 to 662)
  AUTHORS   Gee,S.L. and Conboy,J.G.
  TITLE     Mouse erythroid cells express multiple putative RNA helicase genes
            exhibiting high sequence conservation from yeast to mammals
  JOURNAL   Gene 140 (2), 171-177 (1994)
  MEDLINE   94192995
  REMARK    SEQUENCE FROM N.A.
            TISSUE=ERYTHROLEUKEMIA
COMMENT     ---------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------------.
            [FUNCTION] PUTATIVE ATP-DEPENDENT RNA HELICASE. IT MAY PLAY A ROLE
            IN TRANSLATIONAL ACTIVATION OF MRNA IN THE OOCYTE AND EARLY EMBRYO.
            [TISSUE SPECIFICITY] DEVELOPMENTALLY REGULATED.
            [DEVELOPMENTAL STAGE] EXPRESSED IN OOCYTES. UBIQUITOUSLY FOUND IN 9
            DAYS POST-CONCEPTION EMBRYO, AT LATER STAGES IT IS RESTRICTED TO
            BRAIN AND KIDNEY.
            [SIMILARITY] TO OTHER 'DEAD' BOX FAMILY HELICASES. BELONGS TO THE
            PL10 SUBFAMILY.
FEATURES              Location/Qualifiers
     source           1..662
                      /organism="Mus musculus"
```

```
                     /db_xref="taxon:10090"
     gene            1..662
                     /gene="DDX3"
                     /note="synonyms: DEAD3, ERH, D1PAS1-RS2"
     Protein         1..662
                     /gene="DDX3"
                     /product="DEAD BOX PROTEIN 3"
     Site            224..231
                     /gene="DDX3"
                     /site_type="np-binding"
                     /note="ATP (POTENTIAL)."
     Site            347..350
                     /gene="DDX3"
                     /site_type="unclassified"
                     /note="DEAD BOX."
     Region          582..662
                     /gene="DDX3"
                     /region_name="Domain"
                     /note="GLY/SER-RICH."
     Region          609..616
                     /gene="DDX3"
                     /region_name="Domain"
                     /note="POLY-SER."
     Region          624..630
                     /gene="DDX3"
                     /region_name="Domain"
                     /note="POLY-GLY."
     Region          633..641
                     /gene="DDX3"
                     /region_name="Domain"
                     /note="POLY-GLY."
ORIGIN
        1 mshvavenal gldqqfagld lnssdnqsgg staskgryip phlrnreatk gfydkdssgw
       61 ssskdkdays sfgsrgdsrg kssffgdrgs gsrgrfddrg rgdydgiggr gdrsgfgkfe
      121 rggnsrwcdk sdeddwskpl ppserleqel fsggntginf ekyddipvea tgnncpphie
      181 sfsdvemgei imgnieltry trptpvqkha ipiikekrdl macaqtgsgk taafllpils
      241 qiyadgpgea lramkengry grrkqypisl vlaptrelav qiyeearkfs yrsrvrpcvv
      301 yggaeigqqi rdlergchll vatpgrlvdm mergkigldf ckylvldead rmldmgfepq
      361 irriveqdtm ppkgvrhtmm fsatfpkeiq mlardfldey iflavgrvgs tsenitqkvv
      421 wveeidkrsf lldllnatgk dsltlvfvet kkgadsledf lyhegyacts ihgdrsqrdr
      481 eealhqfrsg kspilvatav aargldisnv khvinfdlps dieeyvhrig rtgrvgnlgl
      541 atsffnerni nitkdlldll veakqevpsw lenmafehhy kgssrgrsks srfsggfgar
      601 dyrqssgass ssfsssrass srsgggghgg srgfggggyg gfynsdgygg nynsqgvdww
      661 gn
//
```

(27)

1: O00571. DEAD-box protein ...[gi:3023628]


```
LOCUS       O00571                  662 aa            linear   PRI 15-JUL-1999
DEFINITION  DEAD BOX PROTEIN 3 (HELICASE-LIKE PROTEIN 2) (HLP2) (DEAD BOX, X
            ISOFORM).
ACCESSION   O00571
VERSION     O00571  GI:3023628
DBSOURCE    swissprot: locus DDX3_HUMAN, accession O00571;
            class: standard.
            extra accessions:O15536,created: Jul 15, 1998.
            sequence updated: Jul 15, 1998.
            annotation updated: Jul 15, 1999.
            xrefs: gi: 2148923, gi: 2148924, gi: 3523149, gi: 3523150, gi:
            2580551, gi: 2580552, gi: 2580549, gi: 2580550
            xrefs (non-sequence databases): MIM 300160, PFAMPF00270,
            PFAMPF00271, PROSITEPS00039
KEYWORDS    Helicase; ATP-binding; RNA-binding; DNA-binding.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 662)
  AUTHORS   CHUNG,J., LEE,S.-G. and SONG,K.
  TITLE     Identification of a human homolog of a putative RNA helicase gene
            (mDEAD3) expressed in mouse erythroid cells
  JOURNAL   Korean J. Biochem. 27, 193-197 (1995)
  REMARK    SEQUENCE FROM N.A.
            TISSUE=LIVER, AND HIPPOCAMPUS
REFERENCE   2  (residues 1 to 662)
  AUTHORS   OWSIANKA,A.M. and PATEL,A.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (~APR-1998)
  REMARK    SEQUENCE FROM N.A.
REFERENCE   3  (residues 1 to 662)
  AUTHORS   Lahn,B.T. and Page,D.C.
  TITLE     Functional coherence of the human Y chromosome
  JOURNAL   Science 278 (5338), 675-680 (1997)
  MEDLINE   98022381
  REMARK    SEQUENCE FROM N.A.
COMMENT     ---------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------------.
            [FUNCTION] PUTATIVE ATP-DEPENDENT RNA HELICASE. INTERACTS
            SPECIFICALLY WITH HEPATITIS C VIRUS CORE PROTEIN RESULTING A CHANGE
            IN INTRACELLULAR LOCATION.
            [SIMILARITY] TO OTHER 'DEAD' BOX FAMILY HELICASES. BELONGS TO THE
            PL10 SUBFAMILY.
FEATURES                Location/Qualifiers
     source             1..662
                        /organism="Homo sapiens"
```

```
                     /db_xref="taxon:9606"
     gene            1..662
                     /gene="DDX3"
                     /note="synonym: DBX"
     Protein         1..662
                     /gene="DDX3"
                     /product="DEAD BOX PROTEIN 3"
     Region          50
                     /gene="DDX3"
                     /region_name="Conflict"
                     /note="K -> R (IN REF. 3)."
     Site            224..231
                     /gene="DDX3"
                     /site_type="np-binding"
                     /note="ATP (POTENTIAL)."
     Site            347..350
                     /gene="DDX3"
                     /site_type="unclassified"
                     /note="DEAD BOX."
     Region          582..662
                     /gene="DDX3"
                     /region_name="Domain"
                     /note="GLY/SER-RICH."
ORIGIN
        1 mshvavenal gldqqfagld lnssdnqsgg staskgryip phlrnreatk gfydkdssgw
       61 ssskdkdays sfgsrsdsrg kssffsdrgs gsrgrfddrg rsdydgigsr gdrsgfgkfe
      121 rggnsrwcdk sdeddwskpl ppserleqel fsggntginf ekyddipvea tgnncpphie
      181 sfsdvemgei imgnieltry trptpvqkha ipiikekrdl macaqtgsgk taafllpils
      241 qiysdgpgea lramkengry grrkqypisl vlaptrelav qiyeearkfs yrsrvrpcvv
      301 yggadigqqi rdlergchll vatpgrlvdm mergkigldf ckylvldead rmldmgfepq
      361 irriveqdtm ppkgvrhtmm fsatfpkeiq mlardfldey iflavgrvgs tsenitqkvv
      421 wveesdkrsf lldllnatgk dsltlvfvet kkgadsledf lyhegyacts ihgdrsqrdr
      481 eealhqfrsg kspilvatav aargldisnv khvinfdlps dieeyvhrig rtgrvgnlgl
      541 atsffnerni nitkdlldll veakqevpsw lenmayehhy kgssrgrsks srfsggfgar
      601 dyrqssgass ssfsssrass srsgggghgs srgfggggyg gfynsdgygg nynsqgvdww
      661 gn
//
```

28

1: P24346. Putative ATP-depe...[gi:113825] BLink, Domains, Links

```
LOCUS       P24346                  697 aa            linear   VRT 01-FEB-1996
DEFINITION  PUTATIVE ATP-DEPENDENT RNA HELICASE AN3.
ACCESSION   P24346
VERSION     P24346  GI:113825
DBSOURCE    swissprot: locus AN3_XENLA, accession P24346;
            class: standard.
            created: Mar 1, 1992.
            sequence updated: Mar 1, 1992.
            annotation updated: Feb 1, 1996.
            xrefs: gi: 65059, gi: 65060, gi: 103989, gi: 345587
            xrefs (non-sequence databases): PFAMPF00270, PFAMPF00271,
            PROSITEPS00039
KEYWORDS    Helicase; ATP-binding; RNA-binding.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
            Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae; Xenopodinae;
            Xenopus.
REFERENCE   1  (residues 1 to 697)
  AUTHORS   Gururajan,R., Perry-O'Keefe,H., Melton,D.A. and Weeks,D.L.
  TITLE     The Xenopus localized messenger RNA An3 may encode an ATP-dependent
            RNA helicase
  JOURNAL   Nature 349 (6311), 717-719 (1991)
  MEDLINE   91141586
  REMARK    SEQUENCE FROM N.A.
COMMENT     ---------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------------.
            [FUNCTION] PUTATIVE ATP-DEPENDENT RNA HELICASE.
            [SIMILARITY] TO OTHER 'DEAD' BOX FAMILY HELICASES. BELONGS TO THE
            PL10 SUBFAMILY.
FEATURES             Location/Qualifiers
     source          1..697
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
     gene            1..697
                     /gene="AN3"
     Protein         1..697
                     /gene="AN3"
                     /product="PUTATIVE ATP-DEPENDENT RNA HELICASE AN3"
     Site            265..272
                     /gene="AN3"
                     /site_type="np-binding"
                     /note="ATP (BY SIMILARITY)."
     Site            388..391
                     /gene="AN3"
                     /site_type="unclassified"
                     /note="DEAD BOX."
```

```
     Region           623..697
                      /gene="AN3"
                      /region_name="Domain"
                      /note="GLY/SER-RICH."
ORIGIN
        1 mshvavenvl nldqqfagld lnsadaesgv agtkgryipp hlrnkeasrn dsnwdsgrgg
       61 ngyingmqdd rdgrmngydr ggygsrgtgr sdrgfydren sgwnsgrdkd ayssfgsrgd
      121 rgkgslfner gsgsrrtddr rqdgfdgmgn rsdksgfgrf drgnsrwsdd rndeddwskp
      181 lapndrveqe lfsgsntgin fekyddipve atgsncpphi esfhdvtmge iimgniqltr
      241 ytrptpvqkh aipiiiekrd lmacaqtgsg ktaafllpil sqiyadgpgd amkhlqengr
      301 ygrrkqfpls lvlaptrela vqiyeearkf ayrsrvrpcv vyggadigqq irdlergchl
      361 lvatpgrlvd mmergkigld fckylvldea drmldmgfep qirriveqdt mppkgvrqtm
      421 mfsatfpkei qilardflde yiflavgrvg stsenitqkv vwveemdkrs flldllnatg
      481 kdsltlvfve tkkgadaled flyhegyact sihgdrsqrd reealhqfrs gkspilvata
      541 vaargldisn vkhvinfdlp sdieeyvhri grtgrvgnlg latsffnekn initkdlldl
      601 lveakqevps wlenmayeqh hksssrgrsk srfsggfgak dyrqssgags sfgssrggrs
      661 sghggsrgfg ggyggfynsd gyggnyggss qvdwwgn
//
```

29

1: P16381. Putative ATP-depe...[gi:130256]


```
LOCUS       P16381                    660 aa            linear   ROD 15-JUL-1999
DEFINITION  PUTATIVE ATP-DEPENDENT RNA HELICASE PL10.
ACCESSION   P16381
VERSION     P16381  GI:130256
DBSOURCE    swissprot: locus PL10_MOUSE, accession P16381;
            class: standard.
            created: Aug 1, 1990.
            sequence updated: Aug 1, 1990.
            annotation updated: Jul 15, 1999.
            xrefs: gi: 200388, gi: 200389, gi: 110038
            xrefs (non-sequence databases): PFAMPF00270, PFAMPF00271,
            PROSITEPS00039
KEYWORDS    Helicase; ATP-binding; RNA-binding; DNA-binding; Spermatogenesis.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (residues 1 to 660)
  AUTHORS   Leroy,P., Alzari,P., Sassoon,D., Wolgemuth,D. and Fellous,M.
  TITLE     The protein encoded by a murine male germ cell-specific transcript
            is a putative ATP-dependent RNA helicase
  JOURNAL   Cell 57 (4), 549-559 (1989)
  MEDLINE   89249320
  REMARK    SEQUENCE FROM N.A.
            TISSUE=TESTIS
COMMENT     ---------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------.
            [FUNCTION] PUTATIVE ATP-DEPENDENT RNA HELICASE. POSSIBLE ROLE IN A
            KEY STEP OF THE SPERMATOGENIC PROCESS.
            [TISSUE SPECIFICITY] TESTIS.
            [DEVELOPMENTAL STAGE] HIGH LEVELS OF PL10 DURING THE MEIOTIC AND
            HAPLOID STAGES OF SPERMATOGENESIS.
            [SIMILARITY] TO OTHER 'DEAD' BOX FAMILY HELICASES. BELONGS TO THE
            PL10 SUBFAMILY.
FEATURES             Location/Qualifiers
     source          1..660
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            1..660
                     /gene="PL10"
                     /note="synonym: D1PAS1"
     Protein         1..660
                     /gene="PL10"
                     /product="PUTATIVE ATP-DEPENDENT RNA HELICASE PL10"
     Site            223..230
                     /gene="PL10"
                     /site_type="np-binding"
```

```
                     /note="ATP (POTENTIAL)."
     Site            346..349
                     /gene="PL10"
                     /site_type="unclassified"
                     /note="DEAD BOX."
     Site            495..514
                     /gene="PL10"
                     /site_type="DNA binding"
                     /note="POTENTIAL."
     Region          581..660
                     /gene="PL10"
                     /region_name="Domain"
                     /note="GLY/SER-RICH."
ORIGIN
        1 mshvaeedel gldqqlagld ltsrdsqsgg staskgryip phlrnreaak afydkdgsrw
       61 skdkdayssf gsrsdtraks sffsdrggsg srgrfdergr sdyesvgsrg grsgfgkfer
      121 ggnsrwcdka deddwskplp pserleqelf sggntginfe kyddipveat gnncpphies
      181 fsdvemgeii mgnieltryt rptpvqkhai piikekrdlm acaqtgsgkt aafllpilsq
      241 iytdgpgeal ramkengkyg rrkqypislv laptrelavq iyeearkfsy rsrvrpcvvy
      301 ggadigqqir dlergchllv atpgrlvdmm ergkigldfc kylvldeadr mldmgfepqi
      361 rriveqdtmp pkgvrhtmmf satfpkeiqm lardfldeyi flavgrvgst senitqkvvw
      421 veeadkrsfl ldllnatgkd slilvfvetk kgadsledfl yhegyactsi hgdrsqrdre
      481 ealhqfrsgk spilvatava argldisnvk hvinfdlpsd ieeyvhrigr tgrvgnlgla
      541 tsffnernin itkdlldllv eakqevpswl enmafehhyk ggsrgrsksr fsggfgardy
      601 rqssgassss fssgrasnsr sgggshgssr gfgggsyggf ynsdgyggny ssqgvdwwgn
//
```

1: O15523. DEAD-box protein ...[gi:6014945]

BLink, Domains, Links

```
LOCUS       O15523                   660 aa            linear   PRI 15-JUL-1999
DEFINITION  DEAD BOX PROTEIN 3, Y-CHROMOSOMAL.
ACCESSION   O15523
VERSION     O15523  GI:6014945
DBSOURCE    swissprot: locus DDXY_HUMAN, accession O15523;
            class: standard.
            created: Jul 15, 1999.
            sequence updated: Jul 15, 1999.
            annotation updated: Jul 15, 1999.
            xrefs: gi: 2580555, gi: 2580556, gi: 2580553, gi: 2580554
            xrefs (non-sequence databases): MIM 400010, PFAMPF00270,
            PFAMPF00271, PROSITEPS00039
KEYWORDS    Helicase; ATP-binding; RNA-binding; DNA-binding.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 660)
  AUTHORS   Lahn,B.T. and Page,D.C.
  TITLE     Functional coherence of the human Y chromosome
  JOURNAL   Science 278 (5338), 675-680 (1997)
  MEDLINE   98022381
  REMARK    SEQUENCE FROM N.A.
COMMENT     ---------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------------.
            [SIMILARITY] TO OTHER 'DEAD' BOX FAMILY HELICASES. BELONGS TO THE
            PL10 SUBFAMILY.
FEATURES             Location/Qualifiers
     source          1..660
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1..660
                     /gene="DBY"
     Protein         1..660
                     /gene="DBY"
                     /product="DEAD BOX PROTEIN 3, Y-CHROMOSOMAL"
     Site            222..229
                     /gene="DBY"
                     /site_type="np-binding"
                     /note="ATP (POTENTIAL)."
     Site            345..348
                     /gene="DBY"
                     /site_type="unclassified"
                     /note="DEAD BOX."
ORIGIN
        1 mshvvvkndp eldqqlanld lnsekqsgga staskgryip phlrnkeask gfhdkdssgw
       61 scskdkdays sfgsrdsrgk pgyfsergsg srgrfddrgr sdydgignre rpgfgrfers
```

```
121 ghsrwcdksv eddwskplpp serleqelfs ggntginfek yddipveatg sncpphienf
181 sdidmgeiim gnieltrytr ptpvqkhaip iikgkrdlva caqtgsgkta afllpilsqi
241 ytdgpgealk avkengrygr rkqypislvl aptrelavqi yeearkfsyr srvrpcvvyg
301 gadigqqird lergchllva tpgrlvdmme rgkigldfck ylvldeadrm ldmgfepqir
361 riveqdtmpp kgvrhtmmfs atfpkeiqml ardfldeyif lavgrvgsts enitqkvvwv
421 edldkrsfll dilgatgsds ltlvfvetkk gadsledfly hegyactsih gdrsqrdree
481 alhqfrsgks pilvatavaa rgldisnvrh vinfdlpsdi eeyvhrigrt grvgnlglat
541 sffneknmni tkdlldllve akqevpswle nmayehhykg gsrgrsksnr fsggfgardy
601 rqssgssssg fgasrgsssr sggggygdsr gfgggggyggf ynsdgyggny nsqgvdwwgn
//
```

1: AL042306. DKFZp434M0720_r1 ...[gi:5421648] Links

IDENTIFIERS

dbEST Id: **2890414**
EST name: DKFZp434M0720_r1
GenBank Acc: AL042306
GenBank gi: 5421648
Database: RZPD Cross Reference: DKFZp434M0720

CLONE INFO
Clone Id: DKFZp434M0720 (5')
DNA type: cDNA

PRIMERS
PolyA Tail: Unknown

SEQUENCE

GAGAACTTGAAGCCACCATGGGAGATGAAGATTGGGAAGCAGAAATCAACCCTCATATGT
CTTCCTATGTTCCCATATTTGAGAAGGATAGGTATTCTGGAGAAAATGGAGACAATTTTA
ACAGGACTCCAGCTTCATCATCAGAAATGGATGATGGACCTTCTCGAAGAGATCATTTCA
TGAAAAGTGGATTTGCCTCTGGGCGGAATTTTGGAAACAGAGATGCTGGTGAGTGTAATA
AGCGAGATAATACATCCACAATGGGTGGTTTTGGAGTTGGAAAGAGTTTTGGAAACAGAG
GTTTTTCAAACAGCAGGTTTGAAGATGGTGATAGCTCTTGTTTCTGGAGAGAGTCTAGTA
ATGACTGCGAAGATAATCCAACACGGAACAGAGGGGTTTTCAAGAAAGGCGGCTATCGAG
ATGGAAATAATTCAGAAGCTTCAGGGCCATACAGAGAGGTGGAGAGGTAGTTTTCCGAGG
TG

Entry Created: Jul 8 1999
Last Updated: Sep 4 2003

COMMENTS

This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by MediGenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project. No s1 sequence available.
This clone (DKFZp434M0720) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de

LIBRARY
Lib Name: 434 (synonym: htes3)
Organism: Homo sapiens
Tissue type: testis
Develop. stage: adult
Lab host: DH10B
Vector: pSport1
R. Site 1: NotI
R. Site 2: SalI

SUBMITTER

lame: MIPS
:nstitution: MIPS
\ddress: Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany

:ITATIONS

:itle: EST (Ottenwaelder, et al.)
\uthors: Ottenwaelder,B., Obermaier,B., Mewes,H.W., Gassenhuber,J., Wiemann,S.
(ear: 1999
3tatus: Unpublished

4AP DATA

Sequence Revision History

Revision history for AL042306

GI	Version	Update Date	Status
5421648	1	Jul 8 1999 7:06 PM	Live

Accession AL042306 was first seen at NCBI on Jul 8 1999 7:06 PM

Entrez

Search for Genes
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

Help|FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences

Check sequence revision history

How to create WWW links to Entrez

LinkOut

Cubby

Related resources

BLAST

Reference sequence project

LocusLink

Clusters of orthologous groups

Protein reviews on the web

Disclaimer | Write to the Help Desk
NCBI | NLM|NIH

1: AA399611. zt93a07.s1 Soares...[gi:2052633] Links

IDENTIFIERS

dbEST Id:	**1040402**
EST name:	zt93a07.s1
GenBank Acc:	AA399611
GenBank gi:	2052633
GDB Id:	5926405

CLONE INFO

Clone Id:	IMAGE:729876 (3')
Source:	IMAGE Consortium, LLNL
DNA type:	cDNA

PRIMERS

Sequencing:	-41m13 fwd. ET from Amersham
PolyA Tail:	Unknown

SEQUENCE

TTTGACATTTAGAATGCTTTAATATTCCCAGTTAACACCATTTGTATCAGTAACTGCAAT
GTTGTAAGTTTTAGCATCTCACATAACTAGTCAGTAAGGATTTTTTTTTTAAGTGTAGGA
GTGAGAATACAAGGACAGGAGCTATGAGAATGTTAAGTTTTATACTTCTGTTAAAAACTC
AAAAATCAAAACTATTTTCTTCTCTGCATCAAAACCACAGACTTGAAGGATGTTTTGGCT
TTAATCCCATGACTCATCATCTACTGGATTGGGAGCTTGTGAAGAAGAAAACCCAGCTGT
GTTCAAAGTGCTCTTGCCCTTTCTGGTATCAACTGATGCAAACACGTTTCCTCTTGTACT
ACCACTGAAGCCAGGAATGTATGTACTAAAGGCAATTTCTTCCAACCATGCAGGAACATC
CTGTTGAGCATCTGTCAATACTTTTACTAGAGGCTGTGCTAAATGGTTATCCGATTCAAG
ATCAAAAAAGGAAATTGCTCTGCCAGTATTCCCACAACGACCAGTACGCCCAATTCGATG
AACATATTCATCAAT

Quality: High quality sequence stops at base: 474

Entry Created:	Apr 28 1997
Last Updated:	May 16 1997

COMMENTS

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand

PUTATIVE ID: Assigned by submitter
TR:G806464 G806464 VASA-LIKE GENE PROTEIN. ;

LIBRARY

Lib Name:	Soares_testis_NHT
Organism:	Homo sapiens
Sex:	male
Lab host:	DH10B
Vector:	pT7T3D-Pac (Pharmacia) with a modified polylinker
R. Site 1:	Not I
R. Site 2:	Eco RI
Description:	1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo

(dT) primer [5'
TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library went
through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name: Wilson RK
Institution: Washington University School of Medicine
Address: 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
E-mail: est@watson.wustl.edu

CITATIONS

Title: WashU-Merck EST Project 1997
Authors: Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost
,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising
,B., White,Y., Wylie,T., Waterston,R., Wilson,R.
Year: 1997
Status: Unpublished

MAP DATA

1: AA398976. zt93a07.r1 Soares...[gi:2052713] Links

IDENTIFIERS

dbEST Id:	**1040482**
EST name:	zt93a07.r1
GenBank Acc:	AA398976
GenBank gi:	2052713
GDB Id:	5926405

CLONE INFO

Clone Id:	IMAGE:729876 (5')
Source:	IMAGE Consortium, LLNL
DNA type:	cDNA

PRIMERS

Sequencing:	-28m13 rev2 ET from Amersham
PolyA Tail:	Unknown

SEQUENCE

ATTGATGAATATGTTCATCGAATTGGGCGTACTGGTCGTTGTGGGAATACTGGCAGACAA
TTTCCTTTTTTGATCTTGAATCGGATAACCATTTAGCACAGCCTCTAGTAAAAGTATTGA
CAGATGCTCAACAGGATGTTCCTGCATGGTTGGAAGAAATTGCCTTTAGTACATACATTC
CTGGCTTCAGTGGTAGTACAAGAGGAAACGTGTTTGCATCAGTTGATACCAGAAAGGGCA
AGAGCACTTTGAACACAGCTGGGTTTTCTTCTTCACAAGCTCCCAATCCAGTAGATGATG
AGTCATGGGATTAAAGCCAAAACATCCTTCAAGTCTGTGGTTTTGATGCAGAGAAGAAAA
TAGTTTTGATTTTTGAGTTTTTAACAGAAGTATAAAACTTAACATTCTCATAGCTCCTGT
CCTTGTATTCTCACTCCTACACTTAAAAAAAAAATCCTTACTGACTAGTTATGTGAGATG
CTAAAACTTAC

Entry Created:	Apr 28 1997
Last Updated:	May 16 1997

COMMENTS

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Putative full length read
The vector to vector length is 554

PUTATIVE ID	Assigned by submitter TR:G286075 G286075 DROSOPHILA VASA HOMOLOGUE ;

LIBRARY

Lib Name:	Soares_testis_NHT
Organism:	Homo sapiens
Sex:	male
Lab host:	DH10B
Vector:	pT7T3D-Pac (Pharmacia) with a modified polylinker
R. Site 1:	Not I
R. Site 2:	Eco RI
Description:	1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5'

TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library went
through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name: Wilson RK
Institution: Washington University School of Medicine
Address: 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
E-mail: est@watson.wustl.edu

CITATIONS

Title: WashU-Merck EST Project 1997
Authors: Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost
,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising
,B., White,Y., Wylie,T., Waterston,R., Wilson,R.
Year: 1997
Status: Unpublished

MAP DATA

34

1: AA383535. EST96928 Testis I...[gi:2035852] Links

IDENTIFIERS

dbEST Id:	**1024641**
EST name:	EST96928
GenBank Acc:	AA383535
GenBank gi:	2035852

CLONE INFO

Clone Id:	(5' end)
Source:	ATCC
Id in host:	187704
DNA type:	cDNA

PRIMERS

Sequencing:	M13 Reverse
PolyA Tail:	Unknown

SEQUENCE

TTATATATGGGGGAACCCAGCTGGGACATTCAATTCGACAAATAGTACAAGGCTGTAATA
TATTATGTGCTACTCCTGGAAGACTGATGGATATCATAGGCAAAGAAAAGATTGGTCTCA
AACAGATCAAATACTTAGTTTTGGATGAAGCTGATCGCATGTTGGATATGGGTTTTGGTC
CAGAAATGAAGAAGTTAATTNNTTGCCCAGGAATGCCATCAAAGGAACAGCGCCAAACCC
TTATGTTCAGTGCAACTTTTCCAGAGGAAATTCAAAGGTTGGCTGCAGAGTTTTTAAAGT
CAAATTATCTGTTTGTTGCTGTTGGACAAGTGGGT

Entry Created:	Apr 21 1997
Last Updated:	Apr 21 1997

COMMENTS

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)

PUTATIVE ID	Assigned by submitter similar to vasa homolog

LIBRARY

Lib Name:	Testis I
Organism:	Homo sapiens
Sex:	male
Organ:	testis
Develop. stage:	adult
Vector:	pBluescript SK-
R. Site 1:	EcoRI
R. Site 2:	XhoI

SUBMITTER

Name:	Kerlavage, AR
Lab:	Bioinformatics
Institution:	The Institute for Genomic Research
Address:	9712 Medical Center Drive, Rockville, MD 20850 USA
Tel:	3018699056

Fax: 3018699423
E-mail: arkerlav@tigr.org

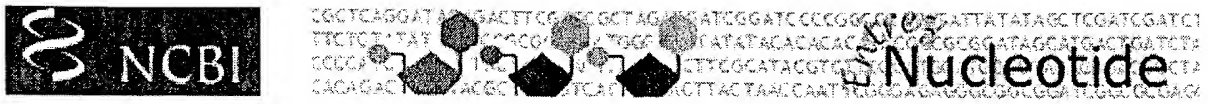
CITATIONS

Medline UID: 96026280
Title: Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
Authors: Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M., Venter,J.C.
Citation: Nature 377 (6547 Suppl): 3-174 1995

MAP DATA

ORIGIN

```
        1 cttgcttcca accctggctc agggcgtcct aaccaggccc ggtagcctct ggggcagact
       61 taggccagag gctgctcaca tgggcctgcc tggggcttga ggtcggcctc tggggccttc
      121 ccgctcagac tggagtcttc atgctgagta gagcggtgtg tgaacacttg aggcctgcag
      181 gggagggagg cccagggaac cttctgctca gcgcctaggc ggccatttct cagaagaaac
      241 ggtgtttgtg gaacctgaag ctatcatggg agatgaagat tgggaggcag aaatcctcaa
      301 acctcatgtg tcttcctacg ttcctgtatt tgagaaggat aaatattctt ctggggcaaa
      361 tggagacact tttaacagga cttcagcttc atcatcagaa atggaagatg gaccttctgg
      421 aagagatcat ttcatgagaa gtggattttc ctctggaaga aatttaggaa acagagatat
      481 tggcgagtct agtaaaagag agactacatc tacaaccggt ggctttggaa gaggaaaggg
      541 ttttggaaac agaggttttt taaataacaa gtttgaagaa ggtgacagct ctggtttctg
      601 gaaagagtct actaatgact gtgaagatac tcagactcgg agcagagggt tttccaagcg
      661 aggcggctat ccagatggga atgattcgga agcttcaggc ccattcagaa gaggtgggag
      721 agatagtgaa tatgaccaag atcagggatc acagcgtggt ggtggccttt ttggttctag
      781 gaaaccagca gcaagtgatt caggcagtgg tgacactttc cagagcagaa gtgggaatgc
      841 ccgaggtgcc tacaaaggct taaatgaaga agtagtaaca ggctctggaa agaattcttg
      901 gaagtcagaa gctgaaggag gcgaaagcag tgatattcaa ggtccaaaag tgacatatat
      961 acccccctcct ccaccagagg atgaggactc catctttgca cattatcaga caggcataaa
     1021 ctttgataaa tatgatacca tacttgttga agtatctgga catgatgcac caccggcaat
     1081 tttgactttt gaagaagcga atctctgcca gaccctgaat aacaacattg ctaaggctgg
     1141 ctataccaag ctcactcctg tgcagaagta cagcattccc attgtgttag caggaagaga
     1201 tttgatggct tgtgctcaaa cagggtctgg gaagacggca gcttttctct tgcctatttt
     1261 ggctcatatg atgagggatg gaataactgc cagtcgcttt aaagaactgc aggaaccaga
     1321 gtgtattatt gtagcaccaa ctcgagaatt gatcaaccaa atttatttgg aagccagaaa
     1381 attttctttt gggacttgtg taagagctgt tgtcatatat ggaggaaccc agtttgggca
     1441 ctcaattcga cagatagtgc aagggtgtaa tatattatgt gctactccag ggaggctgat
     1501 ggatatcata ggcaaagaaa agattggtct caaacaagtc aagtacttag ttttggatga
     1561 agctgatcga atgttggata tgggttttgg acccgaaatg aagaagttaa tttcttgtcc
     1621 aggaatgcca tcaaaggaac agcgccaaac tctcttattc agtgcaactt ttccagaaga
     1681 aatccagagg ttggctggcg agtttttaaa gtcaaattat ttgtttgttg ctgttggaca
     1741 agtgggagga gcttgcagag atgtgcagca gtccattctt caagttggcc cagtattcaa
     1801 aaagagaaaa cttgttgaga ttctacgaaa cataggtgat gaaagaccta tggtctttgt
     1861 tgaaaccaag aaaaaagcag atttcattgc gactttcctt tgtcaagaaa aaatatcaac
     1921 tacaagtatt catggcgatc gggaacagag ggagcgagaa caagctcttg gagattttcg
     1981 ctgtggaaag tgcccagttc ttgttgctac ttcagtggct gccagaggac ttgatattga
     2041 aaatgttcaa catgttatca attttaacct tccttctacc attgatgaat atgttcatcg
     2101 aattggacgt actggacgtt gtggaaatac tggcagggcg atttcttttt ttgataccga
     2161 atctgataat catttagcac aacctctagt taaagtactg tcagatgctc aacaggatgt
     2221 tcctgcgtgg ttagaagaga ttgcctttag ttcatatgcg cctcccagct tcagtaatag
     2281 cacaagaggg gctgtgtttg catcttttga cactaggaag aatttccagg gcaagaacac
     2341 actgaacaca gctgggattt cttctgcaca agctcccaat ccagttgatg acgagtcatg
     2401 ggattaaagc aaacaagcat actgcaagtc tgatggtttt gatgcagaga agaaaaacag
     2461 tttttatttt taaaatttta acagaagtgt gaaacctgat attcttatat ctcctgttct
     2521 tctgttctta ctcccaaccc ttaaaaaata accagcttca ttgattagtt atgcgaaatg
     2581 ctgaagttac aacattgcag ttactgatac aaatggtgtt cactggaaat attaaagcat
     2641 tctatgtttt gcttatttct agtatattct tcagaaagtt aaagacatgt ttcatgtcca
     2701 agtgctatgt cttagtatag tgtttctgat ctataaaaca agcaatagga tatggtgtac
     2761 tcttgtttaa ttatcgggtc taatttctac ttgatccttt aaaagaatag tgtgtcagta
     2821 caatgtatta acatgatttt catgaaacag tggagactga agcctttcaa agttatttga
     2881 tttttagatc atcagacatg taatgaaaat ggttcagttt gcaatgtgag ctctgtactt
     2941 ggtggtatga caaatgtttg cttttataat atacagattt tccttggaaa taaaagatga
     3001 aacacatttc cccctaaaaa aaaaaaaaaa
//
```



1: D14859. Mus musculus mRNA...[gi:286074]


```
LOCUS       MUSDVH                  1930 bp    mRNA    linear   ROD 04-FEB-1999
DEFINITION  Mouse mRNA for drosophila vasa homologue, partial cds.
ACCESSION   D14859
VERSION     D14859.1  GI:286074
KEYWORDS    RNA helicase; drosophila vasa homologue.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1930)
  AUTHORS   Noce,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-APR-1993) Toshiaki Noce, Mitsubishi Kasei Institute
            of Life Sciences, Developmental Biology; 11 Minamiooya, Machida,
            Tokyo 194, Japan (Tel:0427-24-6246, Fax:0427-29-1252)
REFERENCE   2  (bases 1 to 1930)
  AUTHORS   Noce,T.
  JOURNAL   Unpublished (1994)
REFERENCE   3  (sites)
  AUTHORS   Fujiwara,Y., Komiya,T., Kawabata,H., Sato,M., Fujimoto,H.,
            Furusawa,M. and Noce,T.
  TITLE     Isolation of a DEAD-family protein gene that encodes a murine
            homolog of Drosophila vasa and its specific expression in germ cell
            lineage
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91 (25), 12258-12262 (1994)
  MEDLINE   95083681
   PUBMED   7991615
FEATURES             Location/Qualifiers
     source          1..1930
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="BALB/c"
                     /sub_species="domesticus"
                     /db_xref="taxon:10090"
                     /cell_type="primordial germ cell, spermatogonium,
                     spermatocyte"
                     /tissue_type="gonad, testis"
                     /clone_lib="lambda gt10"
                     /dev_stage="adult, embryo"
     CDS             <1..1914
                     /function="RNA helicase"
                     /codon_start=1
                     /product="Drosophila vasa homologue"
                     /protein_id="BAA03584.1"
                     /db_xref="GI:286075"
                     /translation="FGRGKGFGNRGFLNNKFEEGDSSGFWKESNNDCEDNQTRSRGFS
                     KRGGCQDGNDSEASGPFRRGGRGSFRGCRGGFGLGRPNSESDQDQGTQCGGGFLVLGK
                     PAASDSGNGDTYQSRSGSGRGGYKGLNEEVVTGSGKNSWKSETEGGESSDSQGPKVTY
                     IPPPPPEDEDSIFAHYQTGINFDKYDTILVEVSGHDAPPAILTFEEANLCQTLNNNIR
```

IDENTIFIERS

dbEST Id:	**1988110**
EST name:	qf47d11.x1
GenBank Acc:	AI217144
GenBank gi:	3796959

CLONE INFO

Clone Id:	IMAGE:1753173 (3')
Source:	NCI
Insert length:	1013
DNA type:	cDNA

PRIMERS

Sequencing:	-40UP from Gibco
PolyA Tail:	Unknown

SEQUENCE

TTTTTTTTTTTTTTTTTTTTTTTTGACATTTAAAATGCTTTAATATTCCCAGTTAACACC
ATTTGTATCAGTAACTGCAATGTTGTAAGTTTTAGCATCTCACATAACTAGTCAGTAAGG
ATTTTTTTTTTTAAGTGTAGGAGTGAGAATACAAGGACAGGAGCTATGAGAATGTTAAGTT
TTATACTTCTGTTAAAAACTCAAAAATCAAAACTATTTTCTTCTCTGCATCAAAACCACA
GACTTGAAGGATGTTTTGGCTTTAATCCCATGACTCATCATCTACTGGATTGGGAGCTTG
TGAAGAAGAAAACCCAGCTGTGTTCAAAGTGCTCTTGCCCTTTCTGGTATCAACTGATGC
AAACACGTTTCCTCTTGTACTACCACTGAAGCCAGGAATGTATGTACTAAAGGCAATTTC
TTCCAACCATGCAGGAACATCCTGTTGAGCATCTGTCAATACTTTTACTAGAGGCTGTGC
TAAATGGTTATCCGATTCAAGATCAAAAAAGGAAATTGCTCTGCCAGTATTCCCACAACG
ACCAGNACGCCCAAT

Quality: High quality sequence stops at base: 448

Entry Created:	Oct 26 1998
Last Updated:	Nov 10 1998

COMMENTS

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

PUTATIVE ID Assigned by submitter
SW:DDX4_RAT Q64060 DEAD BOX PROTEIN 4 ;

LIBRARY

Lib Name:	Soares_testis_NHT
Organism:	Homo sapiens
Sex:	male
Lab host:	DH10B
Vector:	pT7T3D-Pac (Pharmacia) with a modified polylinker

?. Site 1:	Not I
?. Site 2:	Eco RI
Description:	1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name:	Robert Strausberg, Ph.D.
E-mail:	cgapbs-r@mail.nih.gov

CITATIONS

Title:	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Authors:	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
Year:	1997
Status:	Unpublished

MAP DATA

1: AI953070. wq49h07.x1 NCI_CG...[gi:5745380] Links

IDENTIFIERS

dbEST Id:	**3076455**
EST name:	wq49h07.x1
GenBank Acc:	AI953070
GenBank gi:	5745380

CLONE INFO

Clone Id:	IMAGE:2474653 (3')
Source:	NCI
Insert length:	450
DNA type:	cDNA

PRIMERS

Sequencing:	-40UP from Gibco
PolyA Tail:	Unknown

SEQUENCE

TTTTTTTTTTATGAGAATGTTAAGTTTTATACTTCTGTTAAAAACTCAAAAATCAAAACT
ATTTTCTTCTCTGCATCAAAACCACAGACTTGAAGGATGTTTTGGCTTTAATCCCATGAC
TCATCATCTACTGGATTGGGAGCTTGTGAAGAAGAAAACCCAGCTGTGTTCAAAGTGCTC
TTGCCCTTTCTGGTATCAACTGATGCAAACACGTTTCCTCTTGTACTACCACTGAAGCCA
GGAATGTATGTACTAAAGGCAATTTCTTCCAACCATGCAGGAACATCCTGTTGAGCATCT
GTCAATACTTTTACTAGAGGCTGTGCTAAATGGTTATCCGATTCAAG

Entry Created:	Aug 19 1999
Last Updated:	Mar 8 2000

COMMENTS

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

PUTATIVE ID	Assigned by submitter SW:DDX4_RAT Q64060 DEAD BOX PROTEIN 4 ;

LIBRARY

Lib Name:	NCI_CGAP_GC6
Organism:	Homo sapiens
Tissue type:	pooled germ cell tumors
Lab host:	DH10B
Vector:	pT7T3D-Pac (Pharmacia) with a modified polylinker
R. Site 1:	Not I
R. Site 2:	Eco RI
Description:	Plasmid DNA from the normalized library NCI_CGAP_GC4 was

prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name: Robert Strausberg, Ph.D.
E-mail: cgapbs-r@mail.nih.gov

CITATIONS

Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Authors: NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
Year: 1997
Status: Unpublished

MAP DATA

37

1: AI025074. ov40a04.x1 Soares...[gi:3240687]

IDENTIFIERS

dbEST Id:	**1764280**
EST name:	ov40a04.x1
GenBank Acc:	AI025074
GenBank gi:	3240687

CLONE INFO

Clone Id:	IMAGE:1639758 (3')
Source:	NCI
Insert length:	587
DNA type:	cDNA

PRIMERS

Sequencing:	-40m13 fwd. ET from Amersham
PolyA Tail:	Unknown

SEQUENCE

TTTTGACATTTAGAATGCTTTAATATTCCCAGTTAACACCATTTGTATCAGTAACTGCAA
TGTTGTAAGTTTTAGCATCTCACATAACTAGTCAGTAAGGATTTTTTTTTTAAGTGTAGG
AGTGAGAATACAAGGACAGGAGCTATGAGAATGTTAAGTTTTATACTTCTGTTAAAAACT
CAAAAATCAAAACTATTTTCTTCTCTGCATCAAAACCACAGACTTGAAGGATGTTTTGGC
TTTAATCCCATGACTCATCATCTACTGGATTGGGAGCTTGTGAAGAAGAAAACCCAGCTG
TGTTCAAAGTGCTCTTGCCCTTTCTGGATCAACTGATGCANAACCGTTTCCTCTTGTACT
ACCACTGAAGCCAGGAATGTTGTACTAAAGGCAATTTCTTCCAACCATGCAGGAACATCC
TGTTGAGCATCTGTCAATACTTTACTAGAAGCTGTGCTAAATGGTTATC

Quality:	High quality sequence stops at base: 408
Entry Created:	Aug 13 1998
Last Updated:	Aug 27 1998

COMMENTS

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

PUTATIVE ID	Assigned by submitter SW:DDX4_RAT Q64060 DEAD BOX PROTEIN 4 ;

LIBRARY

Lib Name:	Soares_testis_NHT
Organism:	Homo sapiens
Sex:	male
Lab host:	DH10B
Vector:	pT7T3D-Pac (Pharmacia) with a modified polylinker
R. Site 1:	Not I
R. Site 2:	Eco RI

Description: 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name: Robert Strausberg, Ph.D.
E-mail: cgapbs-r@mail.nih.gov

CITATIONS

Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Authors: NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
Year: 1997
Status: Unpublished

MAP DATA

NCBI

Sequence Revision History

Revision history for AI953070

GI	Version	Update Date	Status
5745380	1	Aug 19 1999 7:22 PM	Live

Accession AI953070 was first seen at NCBI on Aug 19 1999 7:22 PM

Entrez

Search for Genes
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

Help|FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences

Check sequence revision history

How to create WWW links to Entrez

LinkOut

Cubby

Related resources

BLAST

Reference sequence project

LocusLink

Clusters of orthologous groups

Protein reviews on the web

Disclaimer | Write to the Help Desk
NCBI | NLM|NIH

38

1: AI654417. wb31f04.x1 NCI_CG...[gi:4738396] Links

IDENTIFIERS

dbEST Id: **2483061**
EST name: wb31f04.x1
GenBank Acc: AI654417
GenBank gi: 4738396

CLONE INFO
Clone Id: IMAGE:2307295 (3')
Source: NCI
Insert length: 376
DNA type: cDNA

PRIMERS
Sequencing: -40UP from Gibco
PolyA Tail: Unknown

SEQUENCE

AAGTGTAGGTTTGAGAATACAAGGACAGGAGCTATGAGAATGTTAAGTTTTATACTTCTG
TTAAAAACTCAAAAATCAAAACTATTTTCTTCTCTGCATCAAAACCACAGACTTGAAGGA
TGTTTTGGCTTTAATCCCATGACTCATCATCTACTGGATTGGGAGCTTGTGAAGAAGAAA
ACCCAGCTGTGTTCAAAGTGCTCTTGCCCTTTCTGGTATCAACTGATGCAAACACGTTTC
CTCTTGTACTACCACTGAAGCCAGGAATGTATGTACTAAAGGCAATTTCTTCCAACCATG

Entry Created: May 4 1999
Last Updated: Dec 17 1999

COMMENTS

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

PUTATIVE ID Assigned by submitter
SW:DDX4_RAT Q64060 DEAD BOX PROTEIN 4 ;

LIBRARY
Lib Name: NCI_CGAP_GC6
Organism: Homo sapiens
Tissue type: pooled germ cell tumors
Lab host: DH10B
Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker
R. Site 1: Not I
R. Site 2: Eco RI
Description: Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name: Robert Strausberg, Ph.D.
E-mail: cgapbs-r@mail.nih.gov

CITATIONS

Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Authors: NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
Year: 1997
Status: Unpublished

MAP DATA

NCBI

Sequence Revision History

Revision history for AI654417

GI	Version	Update Date	Status
4738396	1	May 4 1999 7:20 PM	Live

Accession AI654417 was first seen at NCBI on May 4 1999 7:20 PM

Entrez

Search for Genes
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

Help|FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences

Check sequence revision history

How to create WWW links to Entrez

LinkOut

Cubby

Related resources

BLAST

Reference sequence project

LocusLink

Clusters of orthologous groups

Protein reviews on the web

Disclaimer | Write to the Help Desk
NCBI | NLM|NIH

1: AI337133. qx83b04.x1 NCI_CG...[gi:4074060] Links

IDENTIFIERS

dbEST Id:	**2120875**
EST name:	qx83b04.x1
GenBank Acc:	AI337133
GenBank gi:	4074060

CLONE INFO

Clone Id:	IMAGE:2009071 (3')
Source:	NCI
DNA type:	cDNA

PRIMERS

Sequencing:	-40UP from Gibco
PolyA Tail:	Unknown

SEQUENCE

AAGTGTAGGAGTGAGAATACAAGGACAGGAGCTATGAGAATGTTAAGTTTTATACTTCTG
TTAAAAACTCAAAAATCAAAACTATTTTCTTCTCTGCATCAAAACCACAGACTTGAAGGA
TGTTTTGGCTTTAATCCCATGACTCATCATCTACTGGATTGGGAGCTTGTGAAGAAGAAA
ACCCAGCTGTGTTCAAAGTGCTCTTGCCCTTTCTGGTATCAACTGATGCAAACACGTTTC
CTCTTGTACTACCACTGAAGCCAGGAATGTATGTACTAAAGGCAATTTCTTCCAACCATG

Entry Created:	Dec 29 1998
Last Updated:	Dec 29 1998

COMMENTS

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html

PUTATIVE ID	Assigned by submitter SW:DDX4_RAT Q64060 DEAD BOX PROTEIN 4 ;

LIBRARY

Lib Name:	NCI_CGAP_GC6
Organism:	Homo sapiens
Tissue type:	pooled germ cell tumors
Lab host:	DH10B
Vector:	pT7T3D-Pac (Pharmacia) with a modified polylinker
R. Site 1:	Not I
R. Site 2:	Eco RI
Description:	Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name: Robert Strausberg, Ph.D.
E-mail: cgapbs-r@mail.nih.gov

CITATIONS

Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Authors: NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
Year: 1997
Status: Unpublished

MAP DATA

1: AA758412. ah66g05.s1 Soares...[gi:2806275] Links

IDENTIFIERS

dbEST Id:	**1480155**
EST name:	ah66g05.s1
GenBank Acc:	AA758412
GenBank gi:	2806275

CLONE INFO

Clone Id:	1320632 (3')
Source:	IMAGE Consortium, LLNL
Insert length:	416
DNA type:	cDNA

PRIMERS

Sequencing:	-40m13 fwd. ET from Amersham
PolyA Tail:	Unknown

SEQUENCE

TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGACATTTAGAATGCTTTAATATTCCCAG
TTAACACCATTTGTATCAGTAACTGCAATGTTGTAAGTTTTAGCATCTCACATAACTAGT
CAGTAAGGATTTTTTTTTTAAGTGTAGGAGTGAGAATACAAGGACAGGAGCTATGAGAAT
GTTAAGTTTTATACTTCTGTTAAAAACTCAAAAATCAAAACTATTTTCTTCTCTGCATCA
AAACCACAGACTTGAAGGATGTTTTGGCTTTAATCCCATGACTCATCATCTACTGGATTG
GGAGCTTGTGAAGAAGAAAACCCAGCTGTGTTCAAAGTGCTCTTGCCCTTTCTGGTATCA
ACTGATGCAAA

Quality: High quality sequence stops at base: 206

Entry Created:	Jan 23 1998
Last Updated:	Dec 29 1998

COMMENTS

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Possible reversed clone: similarity on wrong strand

PUTATIVE ID	Assigned by submitter
	SW:DDX4_RAT Q64060 DEAD BOX PROTEIN 4 ;

LIBRARY

Lib Name:	Soares_testis_NHT
Organism:	Homo sapiens
Sex:	male
Lab host:	DH10B
Vector:	pT7T3D-Pac (Pharmacia) with a modified polylinker
R. Site 1:	Not I
R. Site 2:	Eco RI

Description: 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name: Robert Strausberg, Ph.D.
E-mail: cgapbs-r@mail.nih.gov

CITATIONS

Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Authors: NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
Year: 1997
Status: Unpublished

MAP DATA

41

1: AI969018. wq68d02.x1 NCI_CG...[gi:5765757]

IDENTIFIERS

dbEST Id:	**3095021**
EST name:	wq68d02.x1
GenBank Acc:	AI969018
GenBank gi:	5765757

CLONE INFO

Clone Id:	IMAGE:2476419 (3')
Source:	NCI
Insert length:	429
DNA type:	cDNA

PRIMERS

Sequencing:	-40UP from Gibco
PolyA Tail:	Unknown

SEQUENCE

GAATGTATGTACTAAAGGCAATTTCTTCCAACCATGCAGGAACATCCTGTTGAGCATCTG
TCAATACTTTTACTAGAGGCTGTGCTAAATGGTTATCCGATTCAAGAT

Quality: Trace considered overall poor quality.

Entry Created:	Aug 25 1999
Last Updated:	Oct 20 2000

COMMENTS

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html
Trace considered overall poor quality

PUTATIVE ID Assigned by submitter
SW:DDX4_RAT Q64060 DEAD BOX PROTEIN 4 ;

LIBRARY

Lib Name:	NCI_CGAP_GC6
Organism:	Homo sapiens
Tissue type:	pooled germ cell tumors
Lab host:	DH10B
Vector:	pT7T3D-Pac (Pharmacia) with a modified polylinker
R. Site 1:	Not I
R. Site 2:	Eco RI
Description:	Plasmid DNA from the normalized library NCI_CGAP_GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive

hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (cloneIDs 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name: Robert Strausberg, Ph.D.
E-mail: cgapbs-r@mail.nih.gov

CITATIONS

Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Authors: NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
Year: 1997
Status: Unpublished

MAP DATA

NCBI

Sequence Revision History

Revision history for AI969018

GI	Version	Update Date	Status
5765757	1	Aug 25 1999 7:04 PM	Live

Accession AI969018 was first seen at NCBI on Aug 25 1999 7:04 PM

Entrez

Search for Genes
LocusLink provides curated information for human, fruit fly, mouse, rat, and zebrafish

Help|FAQ

Batch Entrez: Upload a file of GI or accession numbers to retrieve protein or nucleotide sequences

Check sequence revision history

How to create WWW links to Entrez

LinkOut

Cubby

Related resources

BLAST

Reference sequence project

LocusLink

Clusters of orthologous groups

Protein reviews on the web

Disclaimer | Write to the Help Desk
NCBI | NLM|NIH

1: AA400066. zu62a08.s1 Soares...[gi:2053869] Links

IDENTIFIERS

dbEST Id:	**1041638**
EST name:	zu62a08.s1
GenBank Acc:	AA400066
GenBank gi:	2053869
GDB Id:	5929433

CLONE INFO

Clone Id:	IMAGE:742550 (3')
Source:	IMAGE Consortium, LLNL
Insert length:	521
DNA type:	cDNA

PRIMERS

Sequencing:	-41m13 fwd. ET from Amersham
PolyA Tail:	Unknown

SEQUENCE

GAATGTATGTACTATAGGCAATTTCTTCCATCCATGTCGGAACATCCTGTTGAGCATCTG
TCAATACTTTTACTAGAGGCTGTGCTACATGGCTAACCGAATC

Quality: Trace considered overall poor quality.

Entry Created: Dec 20 1996
Last Updated: Nov 9 1997

COMMENTS

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Possible reversed clone: similarity on wrong strand

PUTATIVE ID Assigned by submitter
TR:G806464 G806464 VASA-LIKE GENE PROTEIN. ;

LIBRARY

Lib Name:	Soares_testis_NHT
Organism:	Homo sapiens
Sex:	male
Lab host:	DH10B
Vector:	pT7T3D-Pac (Pharmacia) with a modified polylinker
R. Site 1:	Not I
R. Site 2:	Eco RI
Description:	1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name: Wilson RK
Institution: Washington University School of Medicine
Address: 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
E-mail: est@watson.wustl.edu

CITATIONS

Title: WashU-NCI human EST Project
Authors: Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost ,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra ,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan ,F., Theising,B., White,Y., Wylie,T., Waterston,R., Wilson ,R.
Year: 1997
Status: Unpublished

MAP DATA

A13

1: AA862553. oh44b08.s1 NCI_CG...[gi:2955032] Links

IDENTIFIERS

dbEST Id: **1588408**
EST name: oh44b08.s1
GenBank Acc: AA862553
GenBank gi: 2955032

CLONE INFO
Clone Id: IMAGE:1469463 (3')
Source: NCI
Insert length: 1202
DNA type: cDNA

PRIMERS
Sequencing: -40m13 fwd. ET from Amersham
PolyA Tail: Unknown

SEQUENCE

GAATGTATGTACTAAAGGCAATTTCTTCCAACCATGCAGTGACATCATGTTGAGCATCTG
TCAATACTTTTACTAGATGCTGTCTATAATAGGTATCGGA
Quality: Trace considered overall poor quality.

Entry Created: Mar 4 1998
Last Updated: Aug 24 1998

COMMENTS

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D.,
Michael Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing
Center
Clone distribution: NCI-CGAP clone distribution information
can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Trace considered overall poor quality

PUTATIVE ID Assigned by submitter
SW:DDX4_RAT Q64060 DEAD BOX PROTEIN 4 ;

LIBRARY
Lib Name: NCI_CGAP_GC4
Organism: Homo sapiens
Tissue type: pooled germ cell tumors
Lab host: DH10B
Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker
Description: 1st strand cDNA was prepared from 3 pooled germ cell tumors,
and was then primed with a Not I - oligo(dT) primer.
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library is
normalized. Library was constructed by Bento Soares and M.

Fatima Bonaldo.

SUBMITTER

Name: Robert Strausberg, Ph.D.
E-mail: cgapbs-r@mail.nih.gov

CITATIONS

Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Authors: NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap
Year: 1997
Status: Unpublished

MAP DATA

1: AA401568. zu62a08.r1 Soares...[gi:2053983] Links

IDENTIFIERS

dbEST Id:	**1041752**
EST name:	zu62a08.r1
GenBank Acc:	AA401568
GenBank gi:	2053983
GDB Id:	5929433

CLONE INFO

Clone Id:	IMAGE:742550 (5')
Source:	IMAGE Consortium, LLNL
Insert length:	521
DNA type:	cDNA

PRIMERS

Sequencing:	-28m13 rev2 ET from Amersham
PolyA Tail:	Unknown

SEQUENCE

TTCTACCATTGATGAATATGTTCATCGACTTGGGCGTACTGGTCGTTGTGGGAATACTGG
CAGAGCAAGTTTCCTTTTT

Quality: Trace considered overall poor quality.

Entry Created:	Dec 20 1996
Last Updated:	Nov 9 1997

COMMENTS

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand

PUTATIVE ID Assigned by submitter
TR:G286075 G286075 DROSOPHILA VASA HOMOLOGUE ;

LIBRARY

Lib Name:	Soares_testis_NHT
Organism:	Homo sapiens
Sex:	male
Lab host:	DH10B
Vector:	pT7T3D-Pac (Pharmacia) with a modified polylinker
R. Site 1:	Not I
R. Site 2:	Eco RI
Description:	1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo (dT) primer [5' TGTTACCAATCTGAAGTGGGAGCGGCCGCCCAATTTTTTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization to Cot5, and was

constructed by Bento Soares and M. Fatima Bonaldo.

SUBMITTER

Name: Wilson RK
Institution: Washington University School of Medicine
Address: 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
E-mail: est@watson.wustl.edu

CITATIONS

Title: WashU-NCI human EST Project
Authors: Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost ,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra ,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan ,F., Theising,B., White,Y., Wylie,T., Waterston,R., Wilson ,R.
Year: 1997
Status: Unpublished

MAP DATA

1: AA316798. EST188483 HCC cel...[gi:1969147] Links

IDENTIFIERS

dbEST Id: 958127
EST name: EST188483
GenBank Acc: AA316798
GenBank gi: 1969147

CLONE INFO
Clone Id: (5' end)
Source: ATCC
Id in host: 113555
Other ESTs on clone: THC127206
DNA type: cDNA

PRIMERS
Sequencing: M13 Reverse
PolyA Tail: Unknown

SEQUENCE

GAAAGATTGGATTAGACTTTTGCAAATACTTGGTGTTAGATGAAGCTGATCGGATGTTGG
ATATGGGGTTTGAGCCTCAGATTCGTAGAATAGTCGAACAAGATACTATGCCTCCAAAGG
GTGTCCGCCACACTATGATGTTTAGTGCTACTTTTCCTAAGGAAATACAGATGCTGGCTC
GTGATTTCTTAGATGAATATATCTTCTTGGCTGTAGGAAGAGTTGGCTCTACCTCTGAAA
ACATCACACAGAAAGTAGTTTGGGTGGAAGAATCAGACAAACGGTCATTTCTGCTTGACC
TCCTAAATGCAACAGGCAAGGATTCACTGACCTTAGTGTTTGTGGAGACCAAAAAGGGTG
CAGATTCTCTGGAGGATTTCTTATACCATGAAGGATACGCATGTACCAGCATCCATGGAG
ACCGTTCTCAGAGGGATAGAGAAGAGGCCCTTCAACAGTTCCGCTCAGGGA

Entry Created: Apr 19 1997
Last Updated: Apr 19 1997

COMMENTS

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)

PUTATIVE ID Assigned by submitter
similar to RNA helicase

LIBRARY
Lib Name: HCC cell line (matastasis to liver in mouse) II
Organism: Homo sapiens
Tissue type: colon
Cell type: KM12SM
Cell line: KM12C(HCC)metastasis into mouse (liver)
Vector: pBluescript SK-
R. Site 1: EcoRI
R. Site 2: XhoI

SUBMITTER
Name: Kerlavage, AR
Lab: Bioinformatics

Institution: The Institute for Genomic Research
Address: 9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
E-mail: arkerlav@tigr.org

CITATIONS

Medline UID: 96026280
Title: Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence
Authors: Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A., Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D., White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C., Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D., Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S., Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr., Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M., Moreno-Palanques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M., Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R., Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y., Bednarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J., Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A., He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K., Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Meissner,P.S., Olsen,H., Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M., Dillion,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C., Fraser,C.M., Venter,J.C.
Citation: Nature 377 (6547 Suppl): 3-174 1995

MAP DATA

1: T85890. yd58g05.r1 Soares...[gi:714242] Links

IDENTIFIERS

dbEST Id:	**156011**
EST name:	yd58g05.r1
GenBank Acc:	T85890
GenBank gi:	714242
GDB Id:	468089

CLONE INFO

Clone Id:	IMAGE:112472 (5')
Insert length:	1604
DNA type:	cDNA

PRIMERS

Sequencing:	M13RP1
PolyA Tail:	Unknown

SEQUENCE

TTTTGCAAATACTTGGTGTTAGATGAAGCTGATCGGATGTTGGATATGGGGTTTGAGCCT
CAGATTCGTAGAATAGTCGAACAAGATACTATGCCTCCAAAGGGTGTCCGCCACACTATG
ATGTTTAGTGCTACTTTTCCTAAGGAAATACAGATGCTGGCTCGTGATTTCTTAGGATGA
ATATATCTTCTTGGGCTGTAGGGAAGGAGTTGGGCTCTACCTCTGGAAAACATCACACAG
GAAAGTAGTTGGGGTGGGAAGGANTCAGGACAAACGGGTCATTTCTGGCTTGACCCTCCC
TAAATGGCAACAGGGGCAAGGGATTTCACTTGACCNTTAGGTGTTTTGTGGGGGAGACCC
CAAAAGGGGGTGCCAGGNTTC

Quality: High quality sequence stops at base: 281

Entry Created:	Mar 17 1995
Last Updated:	Mar 17 1995

COMMENTS

Insert Size: 1604
High quality sequence stops: 281 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

PUTATIVE ID Assigned by submitter
SP:PL10_MOUSE P16381 PUTATIVE ATP-DEPENDENT RNA HELICASE ;

LIBRARY

Lib Name:	Soares fetal liver spleen 1NFLS
Organism:	Homo sapiens
Sex:	male
Organ:	Liver and Spleen
Develop. stage:	20 week-post conception fetus
Lab host:	DH10B (ampicillin resistant)
Vector:	pT7T3D (Pharmacia) with a modified polylinker
R. Site 1:	Pac I
R. Site 2:	Eco RI
Description:	1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

SUBMITTER

Name:	Wilson RK
Institution:	Washington University School of Medicine
Address:	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel:	314 286 1800
Fax:	314 286 1810
E-mail:	est@watson.wustl.edu

CITATIONS

Title:	The WashU-Merck EST Project
Authors:	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M. , Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P., Wilson,R.
Year:	1995
Status:	Unpublished

MAP DATA

47

1: T82153. yd95d08.r1 Soares...[gi:705160] Links

IDENTIFIERS

dbEST Id:	**147970**
EST name:	yd95d08.r1
GenBank Acc:	T82153
GenBank gi:	705160
GDB Id:	471600

CLONE INFO

Clone Id:	IMAGE:115983 (5')
Other ESTs on clone:	yd95d08.s1
Insert length:	860
DNA type:	cDNA

PRIMERS

Sequencing:	M13RP1
PolyA Tail:	Unknown

SEQUENCE

TTTTGCAAATACTTGGTGTTAGATGAAGCTGATCGGATGTTGGATATGGGGTTTGAGCCT
CAGATTCGTAGAATAGTCGAACAAGATACTATGCCTCCAAAGGGTGTCCGCCACACTATG
ATGTTTAGTGCTACTTTTCCTAAGGAAATACAGATGCTGGCTCGTGATTTCTTAGATGAA
TATATCTTCTTGGGCTGTAGGGAAGAGTTGGCTCTACCTCTGAAAACATCACACAGAAAG
TAGTTGGGGTGGGAAGGAATCAGACAAACGGTCATTTCTGGCTTGGACCTCCTAAATGGC
AACAGGGCAAGGGTTCACTTGACCTTAGTGTTTTGTTGGGAGACCCAAAAAGGGGTGCCA
G

Quality: High quality sequence stops at base: 269

Entry Created: Mar 10 1995
Last Updated: Mar 15 1995

COMMENTS

Insert Size: 860
High quality sequence stops: 269 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

PUTATIVE ID Assigned by submitter
SP:PL10_MOUSE P16381 PUTATIVE ATP-DEPENDENT RNA HELICASE ;

LIBRARY

Lib Name:	Soares fetal liver spleen 1NFLS
Organism:	Homo sapiens
Sex:	male
Organ:	Liver and Spleen
Develop. stage:	20 week-post conception fetus
Lab host:	DH10B (ampicillin resistant)
Vector:	pT7T3D (Pharmacia) with a modified polylinker
R. Site 1:	Pac I
R. Site 2:	Eco RI
Description:	1st strand cDNA was primed with a Pac I - oligo(dT) primer

[5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo.

SUBMITTER

Name:	Wilson RK
Institution:	Washington University School of Medicine
Address:	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel:	314 286 1800
Fax:	314 286 1810
E-mail:	est@watson.wustl.edu

CITATIONS

Title:	The WashU-Merck EST Project
Authors:	Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M. , Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P., Wilson,R.
Year:	1995
Status:	Unpublished

MAP DATA
